

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: September 5, 2003, 19:09:56 ; search time 12228 Seconds  
(without alignments)

11602.436 Million cell updates/sec

Title: US-10-056-884-1  
3469

```

Perfect score: 3400
Sequence: 1 caagcactgtgtcctaagtgt.....aaaaaaaaaaaaaaaa 3468

```

Scoring table: IDENTITY\_NUC  
Caccor 10 0 Caccor 1 0

Searched: 2888711 begs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

```

minimum DB seq length: 20000000000

```

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: genbba.\*  
2: gb\_ba.\*  
3: gb\_hcg.\*  
4: gb\_in.\*  
5: gb\_on.\*  
6: gb\_ov.\*  
7: gb\_pat.\*  
8: gb\_ph.\*  
9: gb\_pl.\*  
10: gb\_pr.\*  
11: gb\_ro.\*  
12: gb\_sgs.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hun.\*  
18: em\_in.\*  
19: em\_nm.\*  
20: em\_on.\*  
21: em\_oi.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sgs.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_inv.\*  
32: em\_hcg\_oth.\*  
33: em\_hcg\_mus.\*  
34: em\_hcg\_pht.\*  
35: em\_hcg\_pht.\*  
36: em\_hcg\_nam.\*  
37: em\_hcg\_vtl.\*  
38: em\_ey.\*  
39: em\_hgeo\_hum.\*  
40: em\_hgeo\_mus.\*  
41: em\_hgeo\_oth.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3198.4	92.2	5646	9	AB037738	AB037738 Homo sapi
2	2045.4	59.0	18238	9	AC001935	AC001935 Homo sapi
3	2043.4	58.9	184589	9	AC008716	AC008716 Homo sapi
4	2038.6	58.8	98360	9	AC008473	AC008473 Homo sapi
5	1640.8	47.3	2412	6	AX405760	AX405760 Sequence
6	1661.8	33.5	171949	9	AC008652	AC008652 Homo sapi
7	1161.8	33.5	209114	9	AC008383	AC008383 Homo sapi
8	853.2	24.6	135132	2	AC127242	AC127242 Mus muscu
9	853.2	24.6	186417	2	AC114984	AC114984 Mus muscu
10	819.4	23.6	242667	2	AC117867	AC117867 Rat
11	811.8	23.4	230128	10	AC098707	AC098707 Mus muscu
12	807.2	23.2	249703	10	AC113599	AC113599 Rat
13	367.4	10.6	781	10	BC049734	BC049734 Mus muscu
14	319.2	9.2	175059	2	BX323465	BX323465 Danto rer
15	317.6	9.2	200467	2	BX470157	BX470157 Danto rer
16	317.6	9.2	230261	2	BX510095	BX510095 Danto rer
17	286.4	8.3	144319	2	BX511303	BX511303 Danto rer
18	284.4	8.2	124400	2	BXE03406	BXE03406 Danto rer
19	278.4	8.0	93028	2	BX004755	BX004755 Danto rer
20	278.4	8.0	243835	5	AL935104	AL935104 Zebrafish
21	255.4	7.4	174712	10	AL831725	AL831725 Mouse DNA
22	251.6	7.3	1890	2	AL102891	AL102891 Danto rer
23	220.4	6.4	183038	2	AC107770	AC107770 Mus muscu
24	216.4	6.2	246154	2	AC096871	AC096871 Rat
25	216.4	6.2	322972	2	AC122853	AC122853 Rat
26	212.8	6.1	138872	2	AC131951	AC131951 Homo sapi
27	189.6	5.5	243239	2	AC133800	AC133800 Rat
28	188	5.4	472	6	BD109391	BD109391 EST and e
29	167	4.8	3066	9	BC013764	BC013764 Homo sapi
30	167	4.8	109201	9	AC000403	AC000403 Genomic
31	167	4.8	169362	2	AL136440	AL136440 Homo sapi
32	165.4	4.8	251187	2	AL158875	AL158875 Homo sapi
33	159.8	4.6	145979	2	AC102815	AC102815 Mus muscu
34	156.6	4.5	213336	2	AC133844	AC133844 Rat
35	149.8	4.3	675	10	BC049679	BC049679 Mus muscu
36	129.2	3.7	256933	10	AC099771	AC099771 Mus muscu
37	128.2	3.7	194240	2	AC118142	AC118142 Rat
38	127.6	3.7	203720	2	AC093861	AC093861 Homo sapi
39	114.8	3.3	569	5	AY093634	AY093634 Acipenser
40	104.6	3.0	100029	10	AB041174	AB041174 Continuation (4 of
41	84.4	2.4	2093	7	BC051776	BC051776 Danto rer
42	81.8	2.4	1450	17	AF133205	AF133205 Homo sapi
43	81.8	2.4	2689	9	BC032873	BC032873 Homo sapi
44	80.8	2.3	245548	2	BX330662	BX330662 Danto rer
45	79.2	2.3	4582	9	BC037864	BC037864 Homo sapi

## ALIGNMENTS

RESULT 1					PRI 14-MAR-2000
AB037738					
LOCUS	AB037738	5646 bp	mRNA	linear	
DEFINITION	Homo sapiens mRNA for KIAA1317 protein, partial cds.				
ACCESSION	AB037738				
VERSION	AB037738.1	GI:7243014			
KEYWORDS					
SOURCE	Homo sapiens	(human)			
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa;	Chordata; Craniata;	Vertebrata; Euteleostomi;		
	Mammalia; Eutheria;	Primates; Catarrhini; Hominoidea; Homo.			
REFERENCE					
AUTHORS	Nagase,T., Kikuno,R., Ishikawa,K.I., Hirosewa,M. and Ohara,O.				
TITLE	Prediction of the coding sequence of unidentified human genes.				
	Identification of new cDNA clones from brain which				

JOURNAL code for large proteins in vitro  
DNA Res. 7 (1), 65-73 (2000)  
MEDLINE 20181126  
PUBMED 10718198  
REFERENCE 2 (bases 1 to 5646)  
AUTHORS Ohara, O., Nagase, T. and Kikuno, R.  
JOURNAL Direct Submission  
Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute,  
Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba  
292-0812, Japan (E-mail: cdna@info.kazusa.or.jp,  
URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913,  
Fax: +81-438-52-3914)

FEATURES  
source  
1..5646  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="h13117"  
/issue\_type="brain"  
/clone\_lib="PbluescriptII SK plus"  
1..5646  
/gene="KIAA1317"  
/gene="KIAA1317"  
/note="Start codon is not identified."  
/product="KIAA1317 protein"  
/protein\_id="BA9255.1"  
/db\_xref="GI:7243015"  
/translaction="OQKKGTMALSGNCSRYPREQSAVPSPEVELNVGGQYF  
TRHSLTSLPHSLMFMSPKRDVTLANDLAKSRRFFIRDFLRVYLDLDRQV  
LPHPEKRLKREAYFQPLDVLVLTPEIKOSDFCHDFDADQSGDFRIPC  
SLIPADRKWGTITVGRSCTIGREGQADKPRVPLVCGRISLAKVEGSLTNS  
SRDPRAPERYTSRYLKEHLERAFMDLSECGHVAVCNSVTASFYNQYTDKLTWS  
SYTEVYRERSRSPSHCDCCCKGSGDSESGTSCNDLSTSCDSQSAASPORT  
VICGPVTRNIOITLDRIPIKSGPVLIOQGBMRKSDILRTLTSGSRSNMSKKAV  
KELSTSELEKCIODPIKIKIPBPRKHHPMSLEKHYL"

gene  
CDS  
1071..2378  
/gene="KIAA1317"  
/note="Start codon is not identified."  
/codon\_start=1  
/product="KIAA1317 protein"  
/protein\_id="BA9255.1"  
/db\_xref="GI:7243015"

BASE COUNT 1618 a 1163 c 1150 g 1709 t  
ORIGIN

Query Match 92.2%; Score 3198.4; DB 9; Length 5646;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 3225; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

183 ACCAATACGACATCTGATGTAACCTGGGGAATTGGCTGCTGATGTGAGCTTGATGA 242  
760 ACTAAGACGACATCTGATGTAACCTGGGGAATTGGCTGCTGATGTGAGCTTGATGA 819  
243 AGATTGATATGACGAGTGTATATATTTATGAAAGTAGAGCTCACTACATCCACCA 302  
820 AGATTGATATGACGAGTGTATATATTTATGAAAGTAGAGCTCACTACATCCACCA 879  
303 TCCAGGCTTAACTTTTTCAGCATCTTCACTGCTGAGCTCTTATATATTTGAT 362  
880 TCCAGGCTTAACTTTTTCAGCATCTTCACTGCTGAGCTCTTATATATTTGAT 939  
363 TTTCTGGGGGAAAAATCTGGGATAGAGAGAGTATTTTAAATAAGTATGATCTTT 422  
940 TTTCTGGGGGAAAAATCTGGGATAGAGAGAGTATTTTAAATAAGTATGATCTTT 999  
423 TTTCTGGGGGAAAAATCTGGGATAGAGAGAGTATTTTAAATAAGTATGATCTTT 482  
1000 TTTCTGGGGGAAAAATCTGGGATAGAGAGAGTATTTTAAATAAGTATGATCTTT 1059  
483 AATCAAAATAGCAGCAGAGAAAGGGAATGCTCTGATGAGAACTGATGCTT 542  
1060 AATCAAAATAGCAGCAGAGAAAGGGAATGCTCTGATGAGAACTGATGCTT 1119  
543 AATATCTCTGAGAACAGAGGTCCGACAGTTCCCACTCTTCCCTGAGAGTGTAGAGCTGA 602  
1120 AATATCTCTGAGAACAGAGGTCCGACAGTTCCCACTCTTCCCTGAGAGTGTAGAGCTGA 1179  
603 ATGTGGGGGATCAAGTTATTTTACTGCGCATTCACATGATAGCATCCCTCATTTCC 662

1180 ATGTGGGGGATCAAGTTATTTTACTGCGCATTCACATGATAGCATCCCTCATTTCC 1239  
663 TCTGTGAAAAATGTTTTTCCCAAGAGACACGGCTATATGATCTAGCCAGACTCCA 722  
1240 TCTGTGAAAAATGTTTTTCCCAAGAGACACGGCTATATGATCTAGCCAGACTCCA 1299  
723 AGGGAAGTTTTTTCATGACAGATGATATCTTGTTCCTGATATATTTGACATATCCA 782  
1300 AGGGAAGTTTTTTCATGACAGATGATATCTTGTTCCTGATATATTTGACATATCCA 1359  
783 GGGACAGCAGAGTGTCTCTGCTGATATCACTTTCAGAAAAAGAAAGACTGAAAAAGG 842  
1360 GGGACAGCAGAGTGTCTCTGCTGATATCACTTTCAGAAAAAGAAAGACTGAAAAAGG 1419  
843 CTGATATCTTCAGCTCCCAAGATGATCAATCTCTGATATCACTTTCAGAAAAAGAAAG 902  
1420 CTGATATCTTCAGCTCCCAAGATGATCAATCTCTGATATCACTTTCAGAAAAAGAAAG 1479  
903 GCGCAGATGATATCTTCAGCAGATGATCTTGAAGATGCTCCCAAGAAAGCAGACAGAA 962  
1480 GCGCAGATGATATCTTCAGCAGATGATCTTGAAGATGCTCCCAAGAAAGCAGACAGAA 1539  
963 TCTGCCCCCTTCT 1022  
1540 TCTGCCCCCTTCT 1599  
1023 ACAGAGATCTTCAGCTTGGGACAGAGAGGACAGGACATGACCAAGTTTCCGAGTTTC 1082  
1600 ACAGAGATCTTCAGCTTGGGACAGAGAGGACAGGACATGACCAAGTTTCCGAGTTTC 1659  
1083 CCCGATTTTGTGTTGTGGAAGATTTCTCTTGGCAAAAAGATCTTTTGGAAAACTTTGA 1142  
1660 CCCGATTTTGTGTTGTGGAAGATTTCTCTTGGCAAAAAGATCTTTTGGAAAACTTTGA 1719  
1143 ATGAAACAGAGACCTGATGACAGCCCAAGAAAGATACACCTCCAGATTTTATCTCAAT 1202  
1720 ATGAAACAGAGACCTGATGACAGCCCAAGAAAGATACACCTCCAGATTTTATCTCAAT 1779  
1203 TCAAGCAGCTGGAAGAGGCTTTTGTATGTTGTGACAGGTGATGATCCATGAGTGGCTT 1262  
1780 TCAAGCAGCTGGAAGAGGCTTTTGTATGTTGTGACAGGTGATGATCCATGAGTGGCTT 1839  
1263 GTAACATAGGAGACAGATCTTTTCAACAACATATACATGATGATGATGATGATGAT 1322  
1840 GTAACATAGGAGACAGATCTTTTCAACAACATATACATGATGATGATGATGATGAT 1899  
1323 GCTAAGTGAATATGCTTTTCTTACCGTGAAGCTTCCAGATGATGATGATGATGATGAT 1382  
1900 GCTAAGTGAATATGCTTTTCTTACCGTGAAGCTTCCAGATGATGATGATGATGATGAT 1959  
1383 GCTGCTGAAGATGGAAGAGGTGAACAAAGAGGAGAGGCGACGCTTGTGCAATGAC 1442  
1960 GCTGCTGAAGATGGAAGAGGTGAACAAAGAGGAGAGGCGACGCTTGTGCAATGAC 2019  
1443 TCTCAAGTCAAGCTGCGACAGCAGTCTGAGGCAAGCTTCTCCAGAGAGAGGTCATCT 1502  
2020 TCTCAAGTCAAGCTGCGACAGCAGTCTGAGGCAAGCTTCTCCAGAGAGAGGTCATCT 2079  
1503 GTGATCCCGGAGACACCGCAGACCAATCATCGACTCTGAGACGCTCCATCAAGAGGCGC 1562  
2080 GTGATCCCGGAGACACCGCAGACCAATCATCGACTCTGAGACGCTCCATCAAGAGGCGC 2139  
1563 CTGTCCAGCTGATCCCAAGATGATGAGATGAGGAGGAGAAAGCGCTTATCCGGAATTTGA 1622  
2140 CTGTCCAGCTGATCCCAAGATGATGAGATGAGGAGGAGAAAGCGCTTATCCGGAATTTGA 2199  
1623 CTTCAGGCTCCAGAGAAATGGAACATGACAGCAAAAGAAAGCTTTTAAAGAAAGCTT 1682  
2200 CTTCAGGCTCCAGAGAAATGGAACATGACAGCAAAAGAAAGCTTTTAAAGAAAGCTT 2259  
1683 CAATTGAGAGAGCTGAGAAATGATATCCAGATTTTCTTAAAAAAAATTCAGATC 1742

Db	2260	CAATGAGAGAGAGCTGAGAAATGTATCCAGATTTCTTAAAAATCAAAATCCAGATC	2319
Qy	1743	GGTTCTGAGAGAAAAATCTTGGCAATCTGAATCTTTTAAAGAGATCATCTATAAG	1802
Db	2320	GGTTCTGAGAGAAAAATCTTGGCAATCTGAATCTTTTAAAGAGATCATCTATAAG	2379
Qy	1803	GGAGGGCTGGGGCGGGGAAAAAAGAGCATTTTGAATTAACCTCAAAA	1862
Db	2380	GGAGGGCTGGGGCGGGGAAAAAAGAGCATTTTGAATTAACCTCAAAA	2438
Qy	1863	GGAAATCATTTTAAAGAAAAAATACAACTAATGATGCACTTTCTTAACAACAT	1922
Db	2439	GGAAATCATTTTAAAGAAAAAATACAACTAATGATGCACTTTCTTAACAACAT	2498
Qy	1923	AGTCAATTGATATACCTAGCTGCTACTTTACCTAGTTCACCTTAAATGATGCAACAG	1982
Db	2499	AGTCAATTGATATACCTAGCTGCTACTTTACCTAGTTCACCTTAAATGATGCAACAG	2558
Qy	1983	GTAGATTTCTTTCTAGATGTGGAAGTCAAGAAATCTTTTGTAGTATTTGTTGTTA	2042
Db	2559	GTAGATTTCTTTCTAGATGTGGAAGTCAAGAAATCTTTTGTAGTATTTGTTGTTA	2618
Qy	2043	CTTGCTCCATGTGCTAACTATCTTATATATATAGAGCCAGCTAGCTAAAGTAGCTG	2102
Db	2619	CTTGCTCCATGTGCTAACTATCTTATATATATAGAGCCAGCTAGCTAAAGTAGCTG	2678
Qy	2103	AGAGGCTTGGAGAGATTTATCCCAACTGGGTTTTCTCTGATCTTACCTGCTCCCT	2162
Db	2679	AGAGGCTTGGAGAGATTTATCCCAACTGGGTTTTCTCTGATCTTACCTGCTCCCT	2738
Qy	2163	CTTTGAAATGAGGGATATGTAAGAAAAAGATCTGGCCCAATGGCATTAAGTTGAAATTTT	2222
Db	2739	CTTTGAAATGAGGGATATGTAAGAAAAAGATCTGGCCCAATGGCATTAAGTTGAAATTTT	2798
Qy	2223	AAATTTGGTTTTCTTTTGTATATGGGGTGGGGGAAATGGCAATTTATATGACTTTT	2282
Db	2799	AAATTTGGTTTTCTTTTGTATATGGGGTGGGGGAAATGGCAATTTATATGACTTTT	2858
Qy	2283	CACCTAAATCTATATGTCCAGTTTATATGATCCGATATGATGATATTTGTGCACA	2342
Db	2859	CACCTAAATCTATATGTCCAGTTTATATGATCCGATATGATGATATTTGTGCACA	2918
Qy	2343	CAAGCAACAATAAGTATATATATACATGACGACAGATGCGGAGGCTTACGCTCC	2402
Db	2919	CAAGCAACAATAAGTATATATATACATGACGACAGATGCGGAGGCTTACGCTCC	2978
Qy	2403	AAAGGCTGCTCCCTCCGACGAGGCTCTTTAGAAATTTCAATGATGATGACTTC	2462
Db	2979	AAAGGCTGCTCCCTCCGACGAGGCTCTTTAGAAATTTCAATGATGATGACTTC	3038
Qy	2463	TGACTCTTTCTTAAATCTTTTGGGAAGATTTCCAGGCTTTCTTCAACAACCTTTCTA	2522
Db	3039	TGACTCTTTCTTAAATCTTTTGGGAAGATTTCCAGGCTTTCTTCAACAACCTTTCTA	3098
Qy	2523	ACATCAAAATGACTCTCATCAACAATTTGATATCTTATTTGTAATTAATACCTCA	2582
Db	3099	ACATCAAAATGACTCTCATCAACAATTTGATATCTTATTTGTAATTAATACCTCA	3158
Qy	2583	GGCTCCATTTTACTGCTTGTGCTGCTATTAAGAGAGATGAGAGAGCTGTCA	2642
Db	3159	GGCTCCATTTTACTGCTTGTGCTGCTATTAAGAGAGATGAGAGAGCTGTCA	3218
Qy	2643	AAATCTCTTGTGTAAATTAATCAAAATTTGATATCCCAAAATTTTGTGTAATGAC	2702
Db	3219	AAATCTCTTGTGTAAATTAATCAAAATTTGATATCCCAAAATTTTGTGTAATGAC	3278
Qy	2703	TTCACACTGAGGCTCTGACCCGAACTGAATTAACCTTTTCCATGTTTTCACA	2762
Db	3279	TTCACACTGAGGCTCTGACCCGAACTGAATTAACCTTTTCCATGTTTTCACA	3338
Qy	2763	GTCTTACTGCGCAAGTTTAATGATGCTTTCACATATATCAATTAAGTTCTGT	2822
Db	3339	GTCTTACTGCGCAAGTTTAATGATGCTTTCACATATATCAATTAAGTTCTGT	3398

Qy	2823	GTCTCTGATGTTGTGGAACCTAAGGACAAACACAGACTTGAATTAAGGGTCCGGCTT	2882
Db	3399	GTCTCTGATGTTGTGGAACCTAAGGACAAACACAGACTTGAATTAAGGGTCCGGCTT	3458
Qy	2883	TTCGTTGTTTAAAGAAAGTGTATTCACACACAACTAATTAATTTCTAATAAATTT	2942
Db	3459	TTCGTTGTTTAAAGAAAGTGTATTCACACACAACTAATTAATTTCTAATAAATTT	3518
Qy	2943	TAAACTACAAAGCTACATTTTACTGCTTGTAGCCGTTTTGTTGCTTGGATTG	3002
Db	3519	TAAACTACAAAGCTACATTTTACTGCTTGTAGCCGTTTTGTTGCTTGGATTG	3578
Qy	3003	GGCTTGGCTGTGCCATGCTAGAGATTAGCTGTGCTATTTTATGATGCTGTAAACA	3062
Db	3579	GGCTTGGCTGTGCCATGCTAGAGATTAGCTGTGCTATTTTATGATGCTGTAAACA	3638
Qy	3063	CCAAACAAGGTAACGAGCTCCAGAGTTAAGGTTTCAATTTCAATTAAGAACTATCTT	3122
Db	3639	CCAAACAAGGTAACGAGCTCCAGAGTTAAGGTTTCAATTTCAATTAAGAACTATCTT	3698
Qy	3123	TTCATTTACATCTGATGATATGACACAGCCAAAGAAAGTATTAATGATCCGT	3182
Db	3699	TTCATTTACATCTGATGATATGACACAGCCAAAGAAAGTATTAATGATCCGT	3758
Qy	3183	CCATGTAATCTGTAATTTTACTAAGGTAACCAATACCTTTGATAGCTTGTGCTCAAG	3242
Db	3759	CCATGTAATCTGTAATTTTACTAAGGTAACCAATACCTTTGATAGCTTGTGCTCAAG	3818
Qy	3243	AAAGTTAAAGACAGTTTATTTTTCAGCATCTCTCATGATTTCAATGATGTAACAAAA	3302
Db	3819	AAAGTTAAAGACAGTTTATTTTTCAGCATCTCTCATGATTTCAATGATGTAACAAAA	3878
Qy	3303	ATTAATTTGTAATTAATAGTTGTGTGCAAGACCTCTTAATTTTATTTGCTGTGTG	3362
Db	3879	ATTAATTTGTAATTAATAGTTGTGTGCAAGACCTCTTAATTTTATTTGCTGTGTG	3938
Qy	3363	TGCATGTGTATGTATGATCAAGATTAATAAGCAATTTGATTAATAAAAAA	3422
Db	3939	TGCATGTGTATGTATGATCAAGATTAATAAGCAATTTGATTAATAAAAAA	3998
Qy	3423	AAAAAAAAAAAAA 3434	
Db	3999	AAACATGACTAA 4010	

RESULT 2  
AC019335/c 182638 bp DNA linear HTG 07-JUL-2000  
LOCUS Homo sapiens chromosome 5 clone RP11-427K3, WORKING DRAFT SEQUENCE,  
DEFINITION  
AC019335  
AC019335.2 GI:7231064  
VERSION  
HTG; HTGS; PHASE1; HTGS\_DRAFT.  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
1. (bases 1 to 182638)  
Waterston, R.H.  
The sequence of Homo sapiens clone  
UNPUBLISHED  
2. (bases 1 to 182638)  
Waterston, R.H.  
Direct Subcloning  
Submitted (01-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Mar 13, 2000 this sequence version replaced gi:6652510.  
COMMENT  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0427K03  
----- Summary Statistics -----  
Sequencing vector: M13, 87%  
Chemistry: Dye-terminator Big Dye, 13% of reads  
Chemistry: Dye-terminator Big Dye, 13% of reads  
Assembly program: Phrap, version 0.990319  
Consensus quality: 174376 bases at least Q40  
Consensus quality: 176799 bases at least Q30  
Consensus quality: 178323 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert coverage: 4.40 in Q20 bases; sum-of-contigs  
Quality coverage: 4.46 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
-----  
1 1193: contig of 1193 bp in length  
\* 1194 1293: gap of unknown length  
\* 1294 2928: contig of 1635 bp in length  
\* 2929 3028: gap of unknown length  
\* 3029 5126: contig of 2098 bp in length  
\* 5127 5226: gap of unknown length  
\* 5227 6713: contig of 1487 bp in length  
\* 6714 6814: gap of unknown length  
\* 6814 9443: contig of 2630 bp in length  
\* 9444 9544: gap of unknown length  
\* 9544 13186: contig of 3643 bp in length  
\* 13187 13287: gap of unknown length  
\* 13287 18101: contig of 4815 bp in length  
\* 18102 22857: gap of unknown length  
\* 22857 22957: gap of unknown length  
\* 22957 31608: contig of 8652 bp in length  
\* 31609 31708: gap of unknown length  
\* 31709 42773: contig of 10265 bp in length  
\* 42774 50868: gap of unknown length  
\* 50869 50868: gap of unknown length  
\* 50869 59100: contig of 8132 bp in length  
\* 59101 59200: gap of unknown length  
\* 59201 72609: contig of 13409 bp in length  
\* 72610 72710: gap of unknown length  
\* 72710 86964: contig of 14255 bp in length  
\* 86965 87064: gap of unknown length  
\* 87065 104523: contig of 17458 bp in length  
\* 104523 104623: gap of unknown length  
\* 104623 126361: contig of 21738 bp in length  
\* 126361 126460: gap of unknown length  
\* 126461 149770: contig of 23310 bp in length  
\* 149771 149870: gap of unknown length  
\* 149871 182638: contig of 32768 bp in length.  
-----  
FEATURES  
source  
1. 182638  
/organism="Homo sapiens"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-427K3"  
misc\_feature  
1 1193  
/note="assembly\_name:Contig4"  
misc\_feature  
1294 2928  
/note="assembly\_name:Contig5"  
misc\_feature  
3029 5126  
/note="assembly\_name:Contig6"

misc\_feature 5227..6713  
/note="assembly\_name:Contig7"  
misc\_feature 6814..9443  
/note="assembly\_name:Contig8"  
misc\_feature 9544..13186  
/note="assembly\_name:Contig9"  
misc\_feature 13287..18101  
/note="assembly\_name:Contig10"  
misc\_feature 18202..22856  
/note="assembly\_name:Contig11  
vector\_end:SP6  
vector\_end:left"  
misc\_feature 22957..31608  
/note="assembly\_name:Contig12"  
misc\_feature 31709..41973  
/note="assembly\_name:Contig13"  
misc\_feature 42074..50868  
/note="assembly\_name:Contig14"  
misc\_feature 50969..59100  
/note="assembly\_name:Contig15"  
misc\_feature 59201..72609  
/note="assembly\_name:Contig16"  
misc\_feature 72710..86964  
/note="assembly\_name:Contig17  
clone\_end:T7  
vector\_end:right"  
misc\_feature 87065..104522  
/note="assembly\_name:Contig18"  
misc\_feature 104623..126360  
/note="assembly\_name:Contig19"  
misc\_feature 126461..149770  
/note="assembly\_name:Contig20"  
misc\_feature 149871..182638  
/note="assembly\_name:Contig21"  
BASE COUNT 59178 a 34459 c 33797 g 53497 t 1707 others  
ORIGIN  
Query Match 59.0% Score 2045.4 DB 2; Length 182638;  
Best Local Similarity 98.7% Pred. No. 0;  
Matches 2072; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
QY 1336 TGTCTTACCGTACGCTTCCAGATGTCACCTGACGATGTCGTCGCAAGAA 1395  
Db 179790 TTTCTTTTCAGGAGCCCTTCAGATGTCACCTGACGATGTCGTCGCAAGAA 179731  
QY 1396 TGGCAAGGTGACAAAGAGGAGAGCGGACATCTTTCGCAATGACCTCCGACATCTAG 1455  
Db 179730 TGGCAAGGTGACAAAGAGGAGAGCGGACATCTTTCGCAATGACCTCCGACATCTAG 179671  
QY 1456 CTGGACAGCAGCTGAGCGCAGCTCTCCCGAGAGACGATCATCTGTGCTCCCGTGAC 1515  
Db 179670 CTGGACAGCAGCTGAGCGCAGCTCTCCCGAGAGACGATCATCTGTGCTCCCGTGAC 179611  
QY 1516 ACGCGACGACCAATCAGACTGACACGCTCCCATCAAGAAAGGCGCTGTCCAGCTGAT 1575  
Db 179610 ACGCGACGACCAATCAGACTGACACGCTCCCATCAAGAAAGGCGCTGTCCAGCTGAT 179551  
QY 1576 CCAACAGTCAGAGTCGCGCGGAAAGAGCACTTACTCCGATTCTGACTTCAAGGCTCCAG 1635  
Db 179550 CCAACAGTCAGAGTCGCGCGGAAAGAGCACTTACTCCGACTTCAAGGCTCCAG 179491  
QY 1636 GGAATCCAGCAATGAGAGCAAAAGAGCTGTTAAAGAAAGAGCTCTCAATTGAGAGAA 1695  
Db 179490 GGAATCCAGCAATGAGAGCAAAAGAGCTGTTAAAGAAAGAGCTCTCAATTGAGAGAA 179431  
QY 1696 GCTGAGCAAAATGTTTCCAGATTTCTTAAAGAAAGAGCTGTTTCTTCTAGAG 1755  
Db 179430 GCTGAGCAAAATGTTTCCAGATTTCTTAAAGAAAGAGCTGTTTCTTCTAGAG 179371  
QY 1756 AAAACATCTTGGCAATCTGAATCTTAAAGAAAGATATCATCTTAAGAGAGGCGTGGGGG 1815  
Db 179370 AAAACATCTTGGCAATCTGAATCTTAAAGAAAGATATCATCTTAAGAGAGGCGTGGGGG 179311

OY	1816	CGGGGAAAAAAAAAAAAAAAAAGCATTTTGAATTAACTCATAAAGGAATTCATATT	1875
Db	179310	C-GGGAAAAAGAAAAAAAAAAAGCATTTTGAATTAACTCTTAAAGGAATTCATATT	179255
OY	1876	TAAAGGAAAAAAAAATACACTAATGATGCAATTTCTTGAAACCAATATGCCATTTGAT	1935
Db	179251	TAAAGGAAAAAAAAATACACTAATGATGCAATTTCTTGAAACCAATATGCCATTTGAT	179197
OY	1936	ACTACGCGCTACTTTACTACCTGCTCACCCTTAAATGCAAGGATGATTTCTTC	1995
Db	179191	ACTACGCGCTACTTTACTACCTGCTCACCCTTAAATGCAAGGATGATTTCTTC	179133
OY	1996	TGATGTGGAGATACAAAGAAAACCTTTTAACTATTTGTTTGTATTCCTCCCACT	2055
Db	179131	TGATGTGGAGATACAAAGAAAACCTTTTAACTATTTGTTTGTATTCCTCCCACT	179072
OY	2056	GCTAACATCTTATATATATATATGAGACCGACTACGTAAAGTACGTAGAGGCTTTGGGA	2115
Db	179071	GCTAACATCTTATATATATATATGAGACCGACTACGTAAAGTACGTAGAGGCTTTGGGA	179012
OY	2116	GTCAATTTATCCCAACTGGGTTTTTTCTCTCATCCTTTCTACCTCCCTCTTGAATGAG	2175
Db	179011	GTCAATTTATCCCAACTGGGTTTTTTCTCTCATCCTTTCTACCTCCCTCTTGAATGAG	178952
Y	2176	GTATGGTGTGAAAAAGATCTGGCCCAATGCGAATGTGGAAATTTTAAATTTGGATTTT	2235
Db	178951	GTATGGTGTGAAAAAGATCTGGCCCAATGCGAATGTGGAAATTTTAAATTTGGATTTT	178892
OY	2236	CCTTTGTATTTATGGGGTATGGGGGGAATGCGAATTTAATGACTTTTCACTCAATCTAT	2295
Db	178891	CCTTTGTATTTATGGGGTATGGGGGGAATGCGAATTTAATGACTTTTCACTCAATCTAT	178832
OY	2296	ATGTGCGAGTTTATATTTGATCTCCGATGATGATGATTTGTGCAACACAGACAACTAA	2355
Db	178831	ATGTGCGAGTTTATATTTGATCTCCGATGATGATGATTTGTGCAACACAGACAACTAA	178772
OY	2356	GTATGTATATACATGACGCAACGATATCCAGGGCTTGAATCTCCCAAGGCTGTCTC	2415
Db	178771	GTATGTATATACATGACGCAACGATATCCAGGGCTTGAATCTCCCAAGGCTGTCTC	178712
OY	2416	CTGCTCCAGACGCCCTCTCTTGAATATTTCAATGATGATGCTTGAATCTTTCTTA	2475
Db	178711	CTGCTCCAGACGCCCTCTCTTGAATATTTCAATGATGATGCTTGAATCTTTCTTA	178652
OY	2476	AAATTCCTTTGGGAAGATTTCCAGGCTTTCTTCAACAACCTTTCTAATCAATGACT	2535
Db	178651	AAATTCCTTTGGGAAGATTTCCAGGCTTTCTTCAACAACCTTTCTAATCAATGACT	178592
OY	2536	CTCATCATCAACAATATGTATTCCTATATGTGAATTAATACCTCAGGCTCATATTAC	2595
Db	178591	CTCATCATCAACAATATGTATTCCTATATGTGAATTAATACCTCAGGCTCATATTAC	178532
Y	2596	TGCTTTGCTCTTTCTGCAATGAAAGAAAGAGAGAGAGAGTGGTCAACATTCCTTGTG	2655
Db	178531	TGCTTTGCTCTTTCTGCAATGAAAGAGAGAGAGAGAGTGGTCAACATTCCTTGTG	178472
OY	2656	TTAAAAAAATCAAACTTCATATCCCAAAATTTCTGCTAATATGACTCAACATCAGC	2715
Db	178471	TTAAAAAAATCAAACTTCATATCCCAAAATTTCTGCTAATATGACTCAACATCAGC	178412
OY	2716	TTCTCTACCCGAACTGAAATTAACACCTTTCTGCATATTTCAAGTTCTTAATGCCC	2775
Db	178411	TTCTCTACCCGAACTGAAATTAACACCTTTCTGCATATTTCAAGTTCTTAATGCCC	178355
OY	2776	AAGATTTAATGATGGCTTTTGCATATATCCATATTAAGTTCTGTCTCTGTGTGT	2835
Db	178351	AAGATTTAATGATGGCTTTTGCATATATCCATATTAAGTTCTGTGTCTGTGTGT	178292
OY	2836	TGTGGAATTAAGACACACACATGATTTGAATAGAGGTCCGCTTTTGTGTTTAA	2895
Db	178291	TGTGGAATTAAGACACACACATGATTTGAATAGAGGTCCGCTTTTGTGTTTAA	178232
OY	2896	AGAAAGTTTATTCACACACAACCTAATATTTCTTATAAAAATTTTAACTAAGC	2955

Db	178231	AGAAAGTGTATTCACACACAAACCTAATAATTTCTTATBAAAATTTTAACTACAAAGC	178172
QY	2956	TACATTTTACTGCTCTTGAGCCGCTTTTGTTGGCTTGAGGATTCGGGCTTGCTGTG	3015
Db	178171	TACATTTTACTCTCTTGAGCCGCTTTTGTTGGCTTGAGGATTCGGGCTTGCTGTG	178112
QY	3016	CCGATGCGATGATTTAGCTGTGTCATTTTATATGATGTCTGTAACCCAAACAGTAAAC	3075
Db	178111	CCAGTCTAGAGATTTAGCTGTGTCATTTTATATGATGTCTGTAACCCAAACAGTAAAC	178053
QY	3076	TGAAGCTCCAGAGTTAAAGTTTCAGATTTCTPAATGAAACTATCTTTTCAATTAACATCC	3135
Db	178051	TGAAGCTCCAGAGTTAAAGTTTCAGATTTCTPAATGAAACTATCTTTTCAATTAACATCC	177992
QY	3136	TGACTTGTATAGACACAGCCAAAAGAAATGTTTAAATAGCCATCCGCTCCATGTAACTCTG	3195
Db	177991	TGACTTGTATAGACACAGCCAAAAGAAATGTTTAAATAGCCATCCGCTCCATGTAACTCTG	177932
QY	3196	TATTTTACTAAGTATCCCAATAGCTCTTTCATATAGACTGTGCTCTCAAGAAGTTAAAGAC	3255
Db	177931	TATTTTACTAAGTATCCCAATAGCTCTTTCATATAGACTGTGCTCTCAAGAAGTTAAAGAC	177872
QY	3256	CAGTTTATTTTTCAGCACTTCCTCAGATCTTCAGTGGTAAACCAAAAATATTTGTCAAT	3315
Db	177871	CAGTTTATTTTTCAGCACTTCCTCAGATCTTCAGTGGTAAACCAAAAATATTTGTCAAT	177812
QY	3316	TAAATGTGTGTGTGCAAGCACTCCCTAATTTTGTATTTGTGTGTGTGTGTGTAT	3375
Db	177811	TAAATGTGTGTGTGCAAGCACTCCCTAATTTTGTATTTGTGTGTGTGTGTGTAT	177753
QY	3376	GGCTGTCTCAGCTAAATTAAGCAATTTGATATTAATAAAAAAAAAAAAAAAAAAAAAA	3434
Db	177751	GGCTGTCTCAGCTAAATTAAGCAATTTGATATTAATAAAAAAAAAAAAAAAAAAAAAA	177693
RESULT 3			
AC008716/c			
LOCUS	AC008716	184589 bp	DNA
DEFINITION	Home sapiens chromosome 5 clone CTB-85P21, complete sequence.		
ACCESSION	AC008716		
VERSION	AC008716.7	GI:16195190	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 184589)		
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 184589)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission.		
JOURNAL	Submitted (03-APR-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
REFERENCE	3 (bases 1 to 184589)		
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
REFERENCE	4 (bases 1 to 184589)		
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	On Oct 17, 2001 this sequence version replaced gi:8122137. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www.sbgc.stanford.edu Quality: phred Quality >=40 99.5% of Sequence; Estimated Total Number of Errors is 0.9.		



Db	6572	CTGGAGACGCACTGTGAGGCGCAGCTCTCCCAAGAGACGGTCACTGTGGTCCCGTAGC	6513
Oy	1516	ACGCGACGCAACATCTCCAGACTCTGAGACCGTCCCATCAAGAAAGGCGCTGTCACACTGAT	1575
Db	6512	ACGCGACGCAACATCTCCAGACTCTGAGACCGTCCCATCAAGAAAGGCGCTGTCACACTGAT	6453
Oy	1576	CCAACTGTCAGAGATGCGGCGGAGAAAGCAGCTTACTCCGATTTCTGACTTCAGGGCTCAG	1635
Db	6452	CCAACTGTCAGAGATGCGGCGGAGAAAGCAGCTTACTCCGATTTCTGACTTCAGGGCTCAG	6393
Oy	1636	GGAAATGACATGAGCAGCAAAAAAAGCTGTCTAAAGAAAAGCTTCATTTGAGAGAGA	1695
Db	6392	GGAAATGACAAATGAGACCAAAAAAAGCTGTCTAAAGAAAAGCTTCATTTGAGAGAGA	6333
Oy	1696	GCTGAGAAAATGTATCCAGATTTTCTAAAAAAAATTCAGATCGGTTTCTGAGAG	1755
Db	6332	GCTGAGAAAATGTATCCAGATTTTCTAAAAAAAATTCAGATCGGTTTCTGAGAG	6273
Oy	1756	AAAACATCTTGGCACTCGAATCTTTTAAGAGATGATCATTTAAGGAGGCTGGGGG	1815
Db	6272	AAAACATCTTGGCACTCGAATCTTTTAAGAGATGATCATTTAAGGAGGCTGGGGG	6213
Oy	1816	CGGAGAAAAAAGAGATCATTTTGAATTAACCTCATTAAGAAATTCATATTT	1875
Db	6212	CGGG - AAAAGAAAAAAGATCATTTTGAATTAACCTCATTAAGAAATTCATATTT	6155
Oy	1876	TAAAGGAAAAAATACATATATATGATGACATTTCTTGAAACAATATGTCATGATAT	1935
Db	6154	TAAAGGAAAAAATACATATATATGATGACATTTCTTGAAACAATATGTCATGATAT	6095
Oy	1936	ACTATCGCTTACTTTTACCTGATGACCTTAACTGAATGCAAGGGTGAATTTCTTC	1995
Db	6094	ACTATCGCTTACTTTTACCTGATGACCTTAACTGAATGCAAGGGTGAATTTCTTC	6035
Oy	1996	TGATGTGGAAGTACAAAGAAATCTTTTAAGTATTTGTTGTTTACTTCGGCCAGT	2055
Db	6034	TGATGTGGAAGTACAAAGAAATCTTTTAAGTATTTGTTGTTTACTTCGGCCAGT	5975
Oy	2056	GCTAACATCTTATATATATATAGAGCCACTGTAAGAAAGTACTGAGAGGCTTGGGA	2115
Db	5974	GCTAACATCTTATATATATATAGAGCCACTGTAAGAAAGTACTGAGAGGCTTGGGA	5915
Oy	2116	GTCATTTATCCCAACTGGGTTTTTCTCTCATCTCTTCACTCCCTCTTGTAATGAG	2175
Db	5914	GTCATTTATCCCAACTGGGTTTTTCTCTCATCTCTTCACTCCCTCTTGTAATGAG	5855
Oy	2176	GTATGTGAGAAAAAGTCTGGCCCAAGGCAATGATTAAGTTTGAATTTTGTGTTTT	2235
Db	5854	GTATGTGAGAAAAAGTCTGGCCCAAGGCAATGATTAAGTTTGAATTTTGTGTTTT	5795
Oy	2236	CGTTTGTATATGGGGTGTGGGGGGAATGGCAGATTTATATGACTTTCACTCAATCTAT	2295
Db	5794	CGTTTGTATATGGGGTGTGGGGGGAATGGCAGATTTATATGACTTTCACTCAATCTAT	5735
Oy	2296	ATGTGCGCAGTTTATATGACTCCGATACATGAGTATTTGTGCAACAAGCAACAATTA	2355
Db	5734	ATGTGCGCAGTTTATATGACTCCGATACATGAGTATTTGTGCAACAAGCAACAATTA	5675
Oy	2356	GTATGTATATACATAGACGACAGATGCGCAAGGCTGACCTCCCAAGGCTGAGCTC	2415
Db	5674	GTATGTATATACATAGACGACAGATGCGCAAGGCTGACCTCCCAAGGCTGAGCTC	5615
Oy	2416	GTGCTCCGAGAGGCTCTCTTAGAATTTTCAGATGAGTGTGACTCTTCTTA	2475
Db	5614	GTGCTCCGAGAGGCTCTCTTAGAATTTTCAGATGAGTGTGACTCTTCTTA	5555
Oy	2476	AAATCTTTTGGGAAGTTTCCCAAGCTTTCTTCAACAACCTTCTTAACATCAAAATGACT	2535
Db	5554	AAATCTTTTGGGAAGTTTCCCAAGCTTTCTTCAACAACCTTCTTAACATCAAAATGACT	5495
Oy	2536	CTCATCATCAAAATTTGATTTCTTATTTGAAATTAATACCTCAGGCTCAATTTAC	2595

Db 5494 CTCATCATCAACAAATTGATTCCTTATTGGAATTAATACCTTCAGGCTCCATTTAC 5435  
 QY 2596 TCGTTGCTCTTTGCTGCTGATTAAGAGGATGAGAGAGCTGGCTCAAACTTCCTGTG 2855  
 Db 5434 TCGTTGCTCTTTGCTGCTGATTAAGAGGATGAGAGAGCTGGCTCAAACTTCCTGTG 5375  
 QY 2656 TTAATAAATCAACATTCATATCCAAATTTTGTGCTAAATGATCTCAACTGACC 2715  
 Db 5374 TTAATAAATCAACATTCATATCCAAATTTTGTGCTAAATGATCTCAACTGACC 5315  
 QY 2716 TTCTCAACCTGGAATGAATTAATCAACCTTTTCCAAATTTTCCAAAGTTCTTACTGCC 2775  
 Db 5314 TTCTCAACCTGGAATGAATTAATCAACCTTTTCCAAATTTTCCAAAGTTCTTACTGCC 5255  
 QY 2776 ACAGTTTAATGATGAGGCTTTCCAAATTAATCAACATTAAGTTCTGTGTCTGTGTGT 2835  
 Db 5254 ACAGTTTAATGATGAGGCTTTCCAAATTAATCAACATTAAGTTCTGTGTCTGTGTGT 5195  
 QY 2836 TGTGAACTAGAGCAACACAGTACTGTAATGAAGGTCGGGCTTTGTTGTTTGTAG 2895  
 Db 5194 TGTGAACTAGAGCAACACAGTACTGTAATGAAGGTCGGGCTTTGTTGTTTGTAG 5135  
 QY 2896 AGAAGTGTATTCACACACACCTAATATTTCTTAAATTTTAACTAACAAGC 2955  
 Db 5134 AGAAGTGTATTCACACACACCTAATATTTCTTAAATTTTAACTAACAAGC 5075  
 QY 2956 TACATTTTACTGCTGTTGACCGTTTGTGCTTTGGGATTTGGGCTTTGGCTGTG 3015  
 Db 5074 TACATTTTACTGCTGTTGACCGTTTGTGCTTTGGGATTTGGGCTTTGGCTGTG 5015  
 QY 3016 CCCATGCTAGGATTAAGTGTGCTGATTTTAAATGATGCTGTAACAACCAACAGTTAC 3075  
 Db 5014 CCCATGCTAGGATTAAGTGTGCTGATTTTAAATGATGCTGTAACAACCAACAGTTAC 4955  
 QY 3076 TGAAGCTCCAGAGTTAAGTTTCAATTTCTAATTAAGAACTATCTTTTCAATTAATC 3135  
 Db 4954 TGAAGCTCCAGAGTTAAGTTTCAATTTCTAATTAAGAACTATCTTTTCAATTAATC 4895  
 QY 3136 TGAATGTATAGACACAGCCAAAGAACTGTTAATAGCCATCCCTCATGTAATCTGT 3195  
 Db 4894 TGAATGTATAGACACAGCCAAAGAACTGTTAATAGCCATCCCTCATGTAATCTGT 4835  
 QY 3196 TATTTTACTAGTACCAATAGCTTTTCAATGATGCTGTGCTCAAGAGGTTAAAGAC 3255  
 Db 4834 TATTTTACTAGTACCAATAGCTTTTCAATGATGCTGTGCTCAAGAGGTTAAAGAC 4775  
 QY 3256 CAGTTTATTTTCCAGCTTCCCTCAATTCAGTGTGTAACCAAAATTAATTTGCTCAT 3315  
 Db 4774 CAGTTTATTTTCCAGCTTCCCTCAATTCAGTGTGTAACCAAAATTAATTTGCTCAT 4715  
 QY 3316 TAAATGTTGTGTCAGACCTCTTAATTTGTTTATTTGCTGTGTGCTGATGCTGAT 3375  
 Db 4714 TAAATGTTGTGTCAGACCTCTTAATTTGTTTATTTGCTGTGTGCTGATGCTGAT 4655  
 QY 3376 GTGTATCAAGGTAAATTAAGGCAATTTGATTTAAAAAATTTAAAAAATTTAAAAA 3434  
 Db 4654 GTGTATCAAGGTAAATTAAGGCAATTTGATTTCTGTAGAGAAATCAATGACTAA 4596

RESULT 5  
 LOCUS AX405760 2412 bp DNA linear PAT 14-JUN-2002  
 DEFINITION Sequence 175 from Patent WO0222660.  
 ACCESSION AX405760  
 VERSION AX405760.1 GI:21438959  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 TANG, Y. T., LIU, C., ZHOU, P., ASUNDI, V., ZHANG, J., ZHAO, Q. A., REN, F.,  
 XUE, A. J., YANG, Y., WEHLMAN, T. and DRMANEC, R. T.

TITLE Novel nucleic acids and polypeptides  
 JOURNAL Patent: WO 0222660-A.175 21-MAR-2002;  
 HISEQ, INC. (US)  
 FEATURES  
 source Location/Qualifiers  
 1..2412  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /DB\_xref="taxon:9606"  
 CDS  
 1052..2378  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAD34867.1"  
 /DB\_xref="GI:21438960"  
 /translation="MAISGNCSSRYPREGSSAVPNSPEVEELAVGQVYFTHSTLI  
 SIPHSLMWFSPKPDVNDLADSKSREFFINDGFRLYLDLRLROVLPDHPB  
 KGLKREAEVFLQDVLVLTLPEDIKOSPFBCSHDFEDASQSDTRICPSSLPAD  
 RKVGLITVGRSSCTLGRBQDAKRRVRRLIVCGRIISLAKVFGELNESHDPRA  
 PERVTSRFLYLFKFLERAFPMLSCEGFHVACNSSVTASFINOYTDKINVSYTEVF  
 YRPSRMSPSHCDCCKNGKDGKESGTSNDLSTSSCDSSASPOBTVICGVT  
 ROTNIOTLDRPIKKGPIVLIQOSEMRKSDLRILTSGSRESNMSXKKAIVKXKLSIE  
 EELEKCIQDFLKIKIPDFPERKHPQSEILRKHYH."

BASE COUNT 638 a 585 c 551 g 638 t  
 ORIGIN  
 Query Match 47.3%; Score 1640.8; DB 6; Length 2412;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 1645; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 183 ACCAATACGACATCTGATTAATGGAATGAGGCTGCTGATGAGAGCTGATGGA 242  
 Db 760 ACTAAGAGGACATCTGATTAATGGAATGAGGCTGCTGATGAGAGCTGATGGA 819  
 QY 243 AGATTGATTAAGAGGCTGATTAATGGAATGAGGCTGCTGATGAGAGCTGATGGA 302  
 Db 820 AGATTGATTAAGAGGCTGATTAATGGAATGAGGCTGCTGATGAGAGCTGATGGA 879  
 QY 303 TCCAGGTTTAACTATTTTCCAGATCACTTCACTGAGCTCTTATACATTTTAT 362  
 Db 880 TCCAGGTTTAACTATTTTCCAGATCACTTCACTGAGCTCTTATACATTTTAT 939  
 QY 363 TTCTTGGGGGAAAAATAGTGGATTAAGAGAGTCAATTTTAAATGATGATCCTTT 422  
 Db 940 TTCTTGGGGGAAAAATAGTGGATTAAGAGAGTCAATTTTAAATGATGATCCTTT 999  
 QY 423 TCCCTTTCTTAAAGTTGATCCAAAGATTAAGGCTGTGACTCCATTTGATGCACTTTA 482  
 Db 1000 TCCCTTTCTTAAAGTTGATCCAAAGATTAAGGCTGTGACTCCATTTGATGCACTTTA 1059  
 QY 483 AATCAAAATAGCAGCAGAGAAAGGAAAGGAAAGGCTGTGATGGAATCTGTAGCTGT 542  
 Db 1060 AATCAAAATAGCAGCAGAGAAAGGAAAGGAAAGGCTGTGATGGAATCTGTAGCTGT 1119  
 QY 543 ATTATCTCGAAGAACAGGCTCCGAGTTCCCACTCTCCCTGAGGTGTGAGAGCTGA 602  
 Db 1120 ATTATCTCGAAGAACAGGCTCCGAGTTCCCACTCTCCCTGAGGTGTGAGAGCTGA 1179  
 QY 603 ATGTGCGGGGTCAAGTTATTTTACTCCCATTCACATTAATGATTAATCCCTGATCC 662  
 Db 1180 ATGTGCGGGGTCAAGTTATTTTACTCCCATTCACATTAATGATTAATCCCTGATCC 1239  
 QY 663 TCGTGTGAAAATGTTTTCCTCCAAAGAGACAGGCTTAATGATCTAGCAAGAGCTTCA 722  
 Db 1240 TCGTGTGAAAATGTTTTCCTCCAAAGAGACAGGCTTAATGATCTAGCAAGAGCTTCA 1299  
 QY 723 AGGGAAGGTTTTCATTAAGACAGATGATTTCTGTTCCGTTATTTTGTGAGCTATCTCA 782  
 Db 1300 AGGGAAGGTTTTCATTAAGACAGATGATTTCTGTTCCGTTATTTTGTGAGCTATCTCA 1359  
 QY 783 GGGACAGCAGGTGTCTGCTGATCACTTTCCAGAAAAGAGAGACTGAAAAGGGAAG 842  
 Db 1360 GGGACAGCAGGTGTCTGCTGATCACTTTCCAGAAAAGAGAGACTGAAAAGGGAAG 1419  
 QY 843 CTGAATATCTCCAGCTCCAGACTTGATCAATCTGAGCCCGGATGAAATCAAGCAA 902



```

Db      1420  |||||||
          CTTGATATCTTCACAGCTCCAGACTTGGTCAAACTCTGACCCCGCATGAATCAAGGAA 1479
Qy      903  GCCCAGATGAATTTCTGCCACAGTGACTTTGAAGATCCCTCCCAAGAAAGGACACAAAGAA 962
Db      1480  GCCCAGATGAATTTCTGCCACAGTGACTTTGAAGATCCCTCCCAAGAAAGGACACAAAGAA 1539
Qy      963  TCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1022
Db      1540  TCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1599
Qy      1023  ACAGAGATCTGACCTTGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082
Db      1600  ACAGAGATCTGACCTTGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
Qy      1083  CCGGATTTTGGTTTGGTGGAGAGATTTCTTGGGCAAAAGAAAGTCTTTGGAGAGAGAGAG 1142
Db      1660  CCGGATTTTGGTTTGGTGGAGAGATTTCTTGGGCAAAAGAAAGTCTTTGGAGAGAGAGAG 1719
Qy      1143  ATGAAGACAGAGACCTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1202
Db      1720  ATGAAGACAGAGACCTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1779
Qy      1203  TCAAGCACTGGAAGAGCTTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1262
Db      1780  TCAAGCACTGGAAGAGCTTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1839
Qy      1263  GTAACTCATGAGTGAAGAGATCTTTCATCAACCAATATACAGATGACAGATCTGTGCA 1322
Db      1840  GTAACTCATGAGTGAAGAGATCTTTCATCAACCAATATACAGATGACAGATCTGTGCA 1899
Qy      1323  GCTACACTGAATATGCTTCTTACACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382
Db      1900  GCTACACTGAATATGCTTCTTACACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1959
Qy      1383  GCTGCTGCAAGAGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1442
Db      1960  GCTGCTGCAAGAGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2019
Qy      1443  TCTCCACATCTAGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1502
Db      2020  TCTCCACATCTAGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
Qy      1503  GTGATCCCTGAGACAGCCAGACCAACATCTCAAGCTCTGAGAGAGAGAGAGAGAGAGAG 1562
Db      2080  GTGATCCCTGAGACAGCCAGACCAACATCTCAAGCTCTGAGAGAGAGAGAGAGAGAGAG 2139
Qy      1563  CTGTCAAGCTGATCCAAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1622
Db      2140  CTGTCAAGCTGATCCAAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2199
Qy      1623  CTTCAAGCTCTCAGAGAGATGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1682
Db      2200  CTTCAAGCTCTCAGAGAGATGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2259
Qy      1683  CAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1742
Db      2260  CAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2319
Qy      1743  GGTTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1802
Db      2320  GGTTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2379
Qy      1803  GGAAGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1834
Db      2380  GGAAGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2411

```

RESULT 6  
 AC008652/c AC008652 171949 bp DNA linear PRI 31-JUL-2001  
 LOCUS Homo sapiens chromosome 5 clone CTB-18F1, complete sequence.  
 DEFINITION  
 AC008652

```

VERSION AC008652.6 GI:15042788
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
TITLE (bases 1 to 171949)
JOURNAL Doe Joint Genome Institute and Stanford Human Genome Center.
REFERENCE Direct Submission
AUTHORS Unpublished
TITLE 2 (bases 1 to 171949)
JOURNAL Doe Joint Genome Institute.
REFERENCE Direct Submission
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 171949)
AUTHORS Doe Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
REFERENCE Submitted (31-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
AUTHORS Drive, Walnut Creek, CA 94598, USA
JOURNAL On Jul 31, 2001 this sequence version replaced gi:9256021.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
FEATURES
source
1. .171949
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-18F1"
BASE COUNT 54666 a 33484 c 31658 g 52141 t
ORIGIN
Query Match 33.5% Score 1161.8; DB 9; Length 171949;
Best Local Similarity 99.8%; Pred. No. 1.4e-240;
Matches 1163; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 187 ATACGACATCTGAGTAATCTGGGAAATGGGCTGCTGATGAGAGAGAGAGAGAGAG 246
Db 171266 AGACGACATCTGAGTAATCTGGGAAATGGGCTGCTGATGAGAGAGAGAGAGAG 171207
Qy 247 TGAATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306
Db 171206 TGAATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 171147
Qy 307 GGGTTTAACTACTTTTCAAGATCACTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAG 366
Db 171146 GGGTTTAACTACTTTTCAAGATCACTTCACTGAGAGAGAGAGAGAGAGAGAGAG 171087
Qy 367 TGGGGGAAAAATATAGGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
Db 171086 TGGGGGAAAAATATAGGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 171027
Qy 427 TTCTTCAAGAGTATCCAAAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486
Db 171026 TTCTTCAAGAGTATCCAAAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 170967
Qy 487 AAAATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 546
Db 170966 AAAATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 170907
Qy 547 TCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606
Db 170906 TCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 170847
Qy 607 CGGAGGATCAAGTTTATTTTATCTGCAATTCATGATGATGATGATGATGATGATGATGAT 666
Db 170846 CGGAGGATCAAGTTTATTTTATCTGCAATTCATGATGATGATGATGATGATGATGAT 170787

```

```

QY 667 GTGGAATATGTTTCCCAAGAGACACGGCTAATGATCTAGCCAGAGCTCCAAAGG 726
DB 170786 GTGGAATATGTTTCCCAAGAGACACGGCTAATGATCTAGCCAGAGCTCCAAAGG 170727
QY 727 AAGTTTTATGACAGAGATGATCTGTTCCGTTATATCTGAGCTTCTCAAGGA 786
DB 170726 AAGTTTTATGACAGAGATGATCTGTTCCGTTATATCTGAGCTTCTCAAGGA 170667
QY 787 CAGCAGAGTGTCTGCTGCTGATCACTTTCCAGAAAAGAAACATCGAAAAGAACTGA 846
DB 170666 CAGCAGAGTGTCTGCTGCTGATCACTTTCCAGAAAAGAAACATCGAAAAGAACTGA 170607
QY 847 ATACTTCAGCTCCAGACTGTGTCAAACTCTGACCCCGATGAATCAAGCAAGCCC 906
DB 170606 ATACTTCAGCTCCAGACTGTGTCAAACTCTGACCCCGATGAATCAAGCAAGCCC 170547
QY 907 AGATGAATCTGACACAGTGAATCTTGAAGATGCTCCCAAGAAAGCAACAGAAATCTG 966
DB 170546 AGATGAATCTGACACAGTGAATCTTGAAGATGCTCCCAAGAAAGCAACAGAAATCTG 170487
QY 967 CCCCCCTTCTCCCTGCTCCTGCGCAGCAGAGTGGGGTTTCATTAATCTGAGTTACAG 1026
DB 170486 CCCCCCTTCTCCCTGCTCCTGCGCAGCAGAGTGGGGTTTCATTAATCTGAGTTACAG 170427
QY 1027 AGATTCCTGACCTCTGGGAGAGAGACAGCAGATGATCCAGATTCGAGAGTTCCCGC 1086
DB 170426 AGATTCCTGACCTCTGGGAGAGAGAGACAGCAGATGATCCAGATTCGAGAGTTCCCGC 170387
QY 1087 GATTGGTTTGTGGAAGATTCCTTGGCAAAAGATCTTTGAGAAACCTTGAATGA 1146
DB 170386 GATTGGTTTGTGGAAGATTCCTTGGCAAAAGATCTTTGAGAAACCTTGAATGA 170307
QY 1147 AAGCAGAGACCTGATCGAGCCCCGAAAGATACCTCCAGATTTTATCTCAATTCGA 1206
DB 170306 AAGCAGAGACCTGATCGAGCCCCGAAAGATACCTCCAGATTTTATCTCAATTCGA 170247
QY 1207 GCACCTGAAAAGGCTTTTGATATGTTTGTGAGAGTGTGATTCACATGTGCGCTTAA 1266
DB 170246 GCACCTGAAAAGGCTTTTGATATGTTTGTGAGAGTGTGATTCACATGTGCGCTTAA 170187
QY 1267 CTCAATCGGTGACAGATCTTTATCAACCAATATACAGATGACAGATCTGTCAAGTA 1326
DB 170186 CTCAATCGGTGACAGATCTTTATCAACCAATATACAGATGACAGATCTGTCAAGTA 170127
QY 1327 CACTGAATATGCTCTTACCGTAG 1351
DB 170126 CACTGAATATGCTCTTACCGTAG 170102

```

```

COMMENT
On May 1, 2001 this sequence version replaced gi:13699337.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.8.

FEATURES
source
1..209114
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-222022"
64235 t

BASE COUNT 65378 a 39329 c 40172 g 64235 t

ORIGIN
Query Match 33.5%; Score 1161.8; DB 9; Length 209114;
Best Local Similarity 99.8%; Pred. No. 1.4e-240;
Matches 1163; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 187 ATACGACATCTGATGATCTGGGAAATGGCTGCTGCTGATGATGAGAT 246
DB 168231 AGACGACATCTGATGATCTGGGAAATGGCTGCTGCTGATGATGAGAT 168290
QY 247 TGGATATAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 306
DB 168291 TGGATATAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 168350
QY 307 GGGTTTAATCTATCTTTTGAAGATGATGATGATGATGATGATGATGATGAT 366
DB 168351 GGGTTTAATCTATCTTTTGAAGATGATGATGATGATGATGATGATGATGAT 168410
QY 367 TGGGGGAAAAATCTGGATTAAGAGAGTATTTTAAATGAATGATGATGAT 426
DB 168411 TGGGGGAAAAATCTGGATTAAGAGAGTATTTTAAATGAATGATGATGATGAT 168470
QY 427 TTTCTTCAAGATTGATCAAGAGATGATGATGATGATGATGATGATGATGAT 486
DB 168471 TTTCTTCAAGATTGATCAAGAGATGATGATGATGATGATGATGATGATGAT 168530
QY 487 AAAATAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546
DB 168531 AAAATAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 168590
QY 547 TCCTCGAGAAACAAGAGTCCGAGTCCCAATCTCTTCCCTGAGTGTGAGTGAATGT 606
DB 168591 TCCTCGAGAAACAAGAGTCCGAGTCCCAATCTCTTCCCTGAGTGTGAGTGAATGT 168650
QY 607 CGGGGGTCAGATTATTTTATCTGCGCATTCATGATGATGATGATGATGATGAT 666
DB 168651 CGGGGGTCAGATTATTTTATCTGCGCATTCATGATGATGATGATGATGATGAT 168710
QY 667 GTGGAATATGTTTCCCAAGAGACACGGCTAATGATCTAGCCAGAGCTCCAAAGG 726
DB 168711 GTGGAATATGTTTCCCAAGAGACACGGCTAATGATCTAGCCAGAGCTCCAAAGG 168770
QY 727 AAGTTTTATGACAGAGATGATCTGTTCCGTTATATCTGAGCTTCTCAAGGA 786
DB 168771 AAGTTTTATGACAGAGATGATCTGTTCCGTTATATCTGAGCTTCTCAAGGA 168830
QY 787 CAGCAGAGTGTCTGCTGCTGATCACTTTCCAGAAAAGAAACATCGAAAAGAACTGA 846
DB 168831 CAGCAGAGTGTCTGCTGCTGATCACTTTCCAGAAAAGAAACATCGAAAAGAACTGA 168890
QY 847 ATACTTCAGCTCCAGACTGTGTCAAACTCTGACCCCGATGAATCAAGCAAGCCC 906
DB 168891 ATACTTCAGCTCCAGACTGTGTCAAACTCTGACCCCGATGAATCAAGCAAGCCC 168950
QY 907 AGATGAATCTGACACAGTGAATCTTGAAGATGCTCCCAAGAAAGCAACAGAAATCTG 966
DB 168951 AGATGAATCTGACACAGTGAATCTTGAAGATGCTCCCAAGAAAGCAACAGAAATCTG 169010

```

```

RESULT 7
AC008383 209114 bp DNA linear PRI 01-MAY-2001
LOCUS Home sapiens chromosome 5 clone CTC-222022, complete sequence.
DEFINITION AC008383
AC008383
VERSION AC008383.8 GI:13699395
KEYWORDS HTG
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 209114)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 209114)
DOE Joint Genome Institute.
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 209114)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Submitted (01-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL

```

```

QY 967 CCCCCCTTCCTCCCTGCTCCCTGCGCAGCAGAGGGGTTTCATTACTGAGGTTACAG 1026
Db 169011 CCCCCCTTCCTCCCTGCTCCCTGCGCAGCAGAGGGGTTTCATTACTGAGGTTACAG 169070
QY 1027 AGGATCTGCACCTTGGGCGAGAGGAGGAGAGAGTGCAGATTCGAGATTCGAGGTTCCCG 1086
Db 169071 AGGATCTGCACCTTGGGCGAGAGGAGGAGAGAGTGCAGATTCGAGATTCGAGGTTCCCG 169130
QY 1087 GATTTCGTTTGTGGAAGGATTTCTTGGCAAAAGCTTTGGAGAACTTTGAATGA 1146
Db 169131 GATTTCGTTTGTGGAAGGATTTCTTGGCAAAAGCTTTGGAGAACTTTGAATGA 169190
QY 1147 AACGAGAGACCTTGATCGAGCCCAAGAAAGATACACTTCGACATTTTATCTCAATTCAA 1206
Db 169191 AACGAGAGACCTTGATCGAGCCCAAGAAAGATACACTTCGACATTTTATCTCAATTCAA 169250
QY 1207 GCACCTGGAAGGGCTTTTGTATGTTGTGCAGAGTGTGATTCACATGAGTGGCTGTAA 1266
Db 169251 GCACCTGGAAGGGCTTTTGTATGTTGTGCAGAGTGTGATTCACATGAGTGGCTGTAA 169310
QY 1267 CTCATCGGTGACAGCATCTTTCATCAACCAATACAGATGCAAGATCTGTCAAGCTA 1326
Db 169311 CTCATCGGTGACAGCATCTTTCATCAACCAATACAGATGCAAGATCTGTCAAGCTA 169370
QY 1327 CACGGAATATGCTCTTCTACCGTGAG 1351
Db 169371 CACGGAATATGCTCTTCTACCGTGAG 169395

```

## RESULT 8

```

AC127249/c 135132 bp DNA linear HTG 12-JUN-2003
LOCUS AC127249
DEFINITION Mus musculus chromosome UNK clone RP24-475B8, WORKING DRAFT
ACCESSION AC127249
VERSION AC127249.3 GI:31621481
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

## REFERENCE

```

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 135132)
Wilson, R.K.

```

## JOURNAL

```

TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 135132)
McPherson, J.D. and Waterston, R.H.

```

## JOURNAL

```

TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 135132)

```

## JOURNAL

```

TITLE Direct Submission
JOURNAL Submitted (12-JUN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

```

## COMMENT

```

On Jun 12, 2003 this sequence version replaced gi:2186968.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M.BB0475B08
----- Summary Statistics -----
Sequencing vector: Plasmid; 0%
Chemistry: Dye-terminator; 0% of reads
Chemistry: Dye-terminator; Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 133026 bases at least Q40

```

```

Consensus quality: 133293 bases at least Q30
Consensus quality: 133464 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 8787 8886: contig of 8786 bp in length
* 8887 23581: contig of 14695 bp in length
* 23582 43616: gap of unknown length
* 43617 43717: contig of 19935 bp in length
* 43717 135132: contig of 91416 bp in length.
*
* Location/Qualifiers
* 1. 135132
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /chromosome="UNK"
* /clone="RP24-475B8"

```

```

misc_feature 1..8786 assembly_name:Contig21"
misc_feature 8887..23581 assembly_name:Contig22"
misc_feature 23682..43616 assembly_name:Contig23"
misc_feature 43717..135132 assembly_name:Contig24"
BASE COUNT 44115 a 26731 c 25600 g 38364 t 322 others
ORIGIN

```

```

Query Match 24.6%; Score 853.2; DB 2; Length 135132;
Best Local Similarity 70.7%; Pred. No. 7.5e-174;
Matches 1473; Conservative 0; Mismatches 523; Indels 86; Gaps 22;

```

```

QY 1347 GTGAGCTTCCAGATGTCACCTTCACACGCGGATTTGCTGCGCAAGATGCGAAGGTTG 1406
Db 124743 GTGAGCTTCCAGATGTCACCTTCACACGCGGATTTGCTGCGCAAGATGCGAAGGTTG 124684
QY 1407 ACAAGAAGGGGAGGCGGCGCGCTTTCAGATGACCTTCACATCTAGCTGCGAGAGCC 1466
Db 124683 ACA--AAGAGAGGCGGCGCGCTTTCAGATGACCTTCACATCTAGCTGCGAGAGCC 124627
QY 1467 AGTCTGAGGCCAGCTCTTCCGAGAGACGCTATCTGTGTCCGTGACAGCCAGACCA 1526
Db 124626 AGTCTGAGGCCAGCTCTTCCGAGAGACGCTATCTGTGTCCGTGACAGCCAGACCA 124567
QY 1527 ACATCCAGACTCTGACCGTCCATCAAGAGAGGCGCTGTCAGCTGATCCAACTCNG 1586
Db 124566 ACATCCAGACTCTGATGATGCGCATCAAGAGAGGCGCTGTCAGCTGATCCAACTCNG 124507
QY 1587 AGATGCGGCGGAGAAAGCACTTACTCCGATTTCTGACTTTCAGGCTCCAGGGAATCGACA 1646
Db 124506 AGATGCGGCGGAGAAAGCACTTACTCCGATTTCTGACTTTCAGGGAATCGACA 124447
QY 1647 TGAGCAGCAAAAAGAGCTGTTAAAGAAAGAAAGCTCTCAATTGAGAGAGAGCTGAGAAAT 1706
Db 124446 TAAGCAGCAAAAAGAGCTGTTAAAGAAAGAAAGCTCTCAATTGAGAGAGAGCTGAGAAAT 124387
QY 1707 GTATCAGAGATTTCTTAAAGAAAGAAAGCTGTTTCGAGAGAGAGCTGAGAAAT 1766
Db 124386 GTATCAGAGATTTCTTAAAGAAAGAAAGCTGTTTCGAGAGAGAGCTGAGAAAT 124327
QY 1767 GGAATCTGACATTTTAAAGAGATGATCTTAAAGAGAGAGCTGAGAGAGAGAGAGAGAGAG 1826
Db 124326 GGAATCTGACATTTTAAAGAGATGATCTTAAAGAGAGAGAGCTGAGAGAGAGAGAGAGAGAG 124278
QY 1827 AAAAAAAGAGCTATTTTGAATAATTAACCTCATTAAGAAATTCATATTTTAAAGAGAGAG 1886

```

Db	124277	-----GTAGTCGCACTTTGAAATTAACCTCCCAAGAGACATATGTAAAGAAAA	124222
Qy	1887	AATACATATATGATGACATTTCTTAGAACACAAATAGTCATTTAGTACTGCTTA	1946
Db	124221	TA-ACAACTAACGCTCCATTTGTTATGATACAAAT-GTCAATTTGATGACTACTGCTTA	124164
Qy	1947	CTTACCTAGTACCTCTTAACATGTAAATTCACAGGGGTATATTTCTTTCTAGATGTGAA	2006
Db	124163	CTTGGCTAGCTCCCTTAACGTGTAATTCACAGGGGTATATTTCTTTCTAGATGTGAA	124104
Qy	2007	GTACAGAAAACTTTTATATTTATTTG---TTTGTATCTGCTCCATGGCTACTA	2063
Db	124103	CCAGAAAGAGCTTTATATTTGCTTTCTTTATTTATTTGCTCCATGGCTAGAA	124044
Qy	2064	TCTT-ATATATATGAGACGACCTTAAGTAAATAGTACGAGGCTTTGGAGATCTT	2122
Db	124043	TCTTATGATACACAAACACAGCTACGTGATGATGATGATGATGATGATGATGATG	123984
Qy	2123	ATCCCAACGCTGGTTTTT-----CTCATCTCTTAC	2156
Db	123983	ATCCCAACGCTGGTTTTTGGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTG	123924
Qy	2157	CTCCCTCTTTGA--ATGAGGATATGTAGAAAAAGATCTGGCCCAATGCAATATTTG	2214
Db	123923	CTCCCTCCCTTGACCAAGATATGATGAAAGATATATACCCGCTTATGTTAA	123864
Qy	2215	GAATTTTATTTTGGTTTTTCT---TTTGTATATGGGGTT--GGGGGGAATGACAT	2269
Db	123863	GAATTTATCTTTTCCCTTCCCTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTG	123804
Qy	2270	TATATGACTTTTACATCAATCTATATATGTCAGCTTTATATGATCTCCGATGATAG	2329
Db	123803	TGTATGATTTTCTACATAAATCTATATGTCAGCTTTATATGATCTCCGATGATAG	123744
Qy	2330	TATTTGTCAACACAGCA--CAATATATATATATATATATATATATATATATATAT	2388
Db	123743	CGTTTCTGACACAGCAAGTATATATGCTGATATATATATATATATATATATATAT	123684
Qy	2389	GGCTAGACCTCCCAAGGGGTGGTCCGCTCCACAGACCTCTCTTAGAATTTTCA	2448
Db	123683	GGCTAGACAGCAAGAGGGTAAATCTTACTATACACTGCC--CTACAAAGACATTTCA	123625
Qy	2449	GATGATGAGCTTCTGATCTCTTTCTTAAATTTCTTTGGGAATATTTCCACCTTCT	2508
Db	123624	GACAAACAGGCTCTGCTATTTATTTATTTAAACCTCTGAGCAATTTCCACCTCTT	123565
Qy	2509	CACAACTTTCT---AATCAATATGATCTCTCATATCAAAATTTGATTTCTTAT	2563
Db	123564	GGCAGGACAGACTTTCTAAAGCTGAATAGGCCCCATCATCAAAATTTCTTTCTTAT	123505
Qy	2564	TGTAAATTAATACCTCTAGCTCAATTTTACGTGCTTTGCTCTGCAATTAAGA	2623
Db	123504	TTTGAATTAATTAATCCCAAGGCTCTCTGATTTATATGTTCTTTCCCTACATTAAGA	123445
Qy	2624	GGATGAGAGAGCTGTCAAAATCTCTTGTGTAA---AAATCAAACTTCAATATCC	2680
Db	123444	TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	123385
Qy	2681	ACAAAATTTTCTGCTAATATGATGATGATGATGATGATGATGATGATGATGATGATG	2740
Db	123384	ACAAAACAGAG--TGTGAATATGATCTGCACTGAGCTTCTCGGTCTAATTAACAATCA	123326
Qy	2741	CCCTTTTCTGATTTTCTGAGATTTCTTCTGCTCCCAAGATTAATGATGATGATGATGATG	2800
Db	123325	CCATTTCTGATGATCTCAAGATTTTAACTTGCCCAATTTATTAACAGGCTTTGA	123266
Qy	2801	CATTAATCCCAATTAAGTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2860
Db	123265	TATTAATCCCAAGCAATTTCTGTGCTCTCATATGATGATGATGATGATGATGATGATG	123206
Qy	2861	ACTTAATTAAGGCTCGGCTTTGTT--TGTTTAGAAAGTTGATTTCCACACACAC	2919
Db	123205	GATTGAATATGATCCAGCTTCCTGCTTCTGTTTATAGAGATTTGCCCCAACATAT	123146
Qy	2920	CTAATAATTTCTTATTAATAATTTTAACTAACAGTACATTTTACTCTGTAACCG	2979
Db	123145	GTCATAATTTCTATACAACTTTTACATGCTACCTGTTTACTGTTGCTGTTGATGCA	123086
Qy	2980	TTTTTGTGCTCTTTGGGATTC--GGGCTTTGGCTGTGCCATGATGATTTAGCTGT	3038
Db	123085	GGTTGGATTAATCTTTGGAATCTCGGGGTTGGCTGTGGCCCTACATCGTTAGTGTAT	123026
Qy	3039	CATTTTATATGATCTGTAACACCAAGAGTAACTGAAGCTCAGATTAAGTTTC	3098
Db	123025	AATTTTACAGTGTCTGTAATGACCAAGGGGTGCTGGAACATTAAGTACTAATTT	122966
Qy	3099	AGATTTCTTAATGAATCATCTTTTCAATTAATGATCTGATGATGATGATGATGATGATG	3158
Db	122965	GATTTTATA-----CTTTAAAAATTTCTGACCTGTGTGATGATGATGATGATGATG	122915
Qy	3159	AAGAACTGTATAAGCCATCCGTCATGATGATGATGATGATGATGATGATGATGATG	3218
Db	122914	AAGAACTGTATAAGCCATCCGTCATGATGATGATGATGATGATGATGATGATGATG	122856
Qy	3219	TCTTTCAATGATCTGTGCTACAGAGAGTTTAAAGCCAGTTT--ATTTCAGATTTCT	3277
Db	122855	TCTTTGTGTTAC--TGTGCTAATAAGATGTCCAAAGACATTTTATTTTATGATTTCT	122797
Qy	3278	CATGCAATTCAGTGTATACCAAAATTAATTTGTCATTAATGATTTGTGTCAGACAT	3337
Db	122796	CATGCAATTCAGTGTATACCAAAATTAATTTGTCATTAATGATTTATATACATA---GTGTGTCACAAAGACT	122741
Qy	3338	CTTAATTTGTTTATTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	3379
Db	122740	CCATATTTTGTGTTGT	122699
RESULT 9			
AC114984			
LOCUS	186417 bp	DNA	linear HTG 05-JUN-2003
DEFINITION	Mus musculus clone RP23-248F9, *** SEQUENCING IN PROGRESS ***		6
ACCESSION	AC114984		
VERSION	GI:30984634		
KEYWORDS	HTG, HTGS PHASE1, HTGS FULLTOP, HTGS_ACTIVEFIN.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (Baee 1 to 186417) and Lander E.		
AUTHORS	Birren, B., Nussbaum, C. and Lander E.		
JOURNAL	Mus musculus, clone RP23-248F9		
AUTHORS	Unpublished		
	2 (baee 1 to 186417)		
	Birren, B., Linton, L., Nussbaum, C., Lander, E., All, A., Allen, N.,		
	Anderson, S., Barta, N., Bastien, V., Bloom, T., Boguslavsky, L.,		
	Bouhassal, B., Brown, A., Camarero, J., Campopiano, A., Chang, J.,		
	Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collamore, A.,		
	Cook, A., Cooke, P., DeArlelano, K., Dewar, K., Diaz, D., Dodge, S.,		
	Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, D., Gaidyna, S.,		
	Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,		
	Hagoo, B., Horton, L., Huie, W., Iliev, J., Johnson, R., Jones, C.,		
	Kamut, A., Karatas, A., Kelle, C., Labrecque, K., Lamazares, R.,		
	Lander, E., Lehoucq, J., Levine, R., Lindblad-Toh, K., Liu, G.,		
	Maclean, C., Macdonald, P., Major, J., Marquardt, N., Matthews, C.,		
	McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,		
	Mihova, T., Mienga, V., Murphy, T., O'Donnell, P., O'Neill, D.,		
	Nolte, J., Norman, C.H., O'Connor, T., O'Neill, J., Nguyen, C., Nicol, R.,		
	Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Pollara, V.,		
	Ramond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P.,		
	Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,		
	Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,		
	Struam, K., Subramanian, A., Talama, J., Tesfaye, S., Theodore, J.,		
	Toplam, K., Travers, M., Travis, N., Trigilio, J., Vasilev, H.,		
	Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,		

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (14-MAR-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 186417)

Birtten, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, S., Atchuch, H.M., Barna, N., Bascien, V., Bloom, T.,  
Boguslavsky, L., Bouckhalter, B., Camarata, J., Chang, D., Chepel, Y.,  
Collymore, A., Cook, A., Cooke, P., Corbin, B., Deatellano, K.,  
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galsagan, J., Gadya, S.,  
Graham, L., Grand-Pierre, N., Hafez, N., Hagan, D., Hagedorn, B.,  
Hall, J., Horton, L., Hulme, W., Iller, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Lander, E., Levine, R.,  
Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,  
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Medlitz, J., Menais, L., Minova, I., Mlenka, V., Murphy, T., Naylor, J.,  
O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Piere, N.,  
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Steange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
Vasiliiev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (05-JUN-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 22, 2003 this sequence version replaced gi:30023906.  
All repeats were identified using RepeatMasker:  
Smith, A.P.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: L19035  
Center clone name: 248\_P\_9

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 32842: contig of 32842 bp in length  
\* 32843 32842: gap of 100 bp  
\* 32943 61950: contig of 29008 bp in length  
\* 61951 62050: gap of 100 bp  
\* 62051 70963: contig of 8913 bp in length  
\* 70964 71063: gap of 100 bp  
\* 71064 74030: contig of 2967 bp in length  
\* 74031 74130: gap of 100 bp  
\* 74131 162809: contig of 88679 bp in length  
\* 162810 162909: gap of 100 bp  
\* 162910 186417: contig of 23508 bp in length.

Location/Qualifiers  
1. 186417  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-248P9"  
/clone\_1kb="RP23-248P9"  
/clone\_1kb="RP23-248P9"

BASE COUNT 54373 a 36475 c 36595 g 58359 t 615 others

ORIGIN

Query Match 24.6% Score 853.2; DB 2; Length 186417;  
Best Local Similarity 70.7%; Freq. No. 7.5e-174;  
Matches 1473; Conservative 0; Mismatches 523; Indels 86; Gaps 22;

QY 1347 GTAGCCTTCCAGATGTGACCTTCACTGCGATTGCTGCAAGATGCAAGG 1406  
DB 172660 GTAGCCTTCCAGATGTGACCTTCACTGCGATTGCTGCAAGATGCAAGG 172719

QY 1407 ACAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1466  
DB 172720 ACA---AAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 172776

QY 1467 AGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1526  
DB 172777 AGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 172836

QY 1527 ACATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1586  
DB 172837 ACATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 172896

QY 1587 AGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1646  
DB 172897 AGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 172956

QY 1647 TGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1706  
DB 172957 TGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173016

QY 1707 GTATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1766  
DB 173017 GTATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173076

QY 1767 GGCAGATCTGAACTTTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1826  
DB 173077 GGCAGATCTGAACTTTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173125

QY 1827 AAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1886  
DB 173126 AAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173181

QY 1887 AATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1946  
DB 173182 AATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173239

QY 1947 CTTTACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2006  
DB 173240 CTTTACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173299

QY 2007 GTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2063  
DB 173300 GTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173359

QY 2064 TCTT-AT 2122  
DB 173360 TCTT-AT 173419

QY 2123 ATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2186  
DB 173420 ATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173479

QY 2157 CTCCTCTCTTGA-ATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2214  
DB 173480 CTCCTCTCTTGA-ATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173539

QY 2215 GAATTTTAT 2269  
DB 173540 GAATTTTAT 173599

QY 2270 TTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2329  
DB 173600 TTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173659

QY 2330 TATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2388  
DB 173660 TATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 173719

```

Qy 2389 GGCCTAGACCTCCCAAGGGCTGCTGCTCCCTCCCAAGACCCCTCTTAGAATATTTC 2448
    |||
Db 173720 GGGCCTGACAGAGAGGGCTTAATCTTATCTACCTGCCC-CTACAGAGACCTTCA 173778
Qy 2449 GATGATGAGCTCTCTACCTCTTTCTTAATCTTTTGGAGATATCCAGCCCTTCTT 2508
    |||
Db 173779 GACAAACAGAGCTCTCTATTTATTTAAACCTTCGGGAGATTTCCAGCCCTTCTT 173838
Qy 2509 CACAACACTTCT- - - - -ACATCAATAGACTCATCATCAACAAATGATTCCTAT 2563
    |||
Db 173839 GGCAGGAGACCTTCTTAAGCTGAATAGGCCCATCATCAACAAATTCCTTCTTAT 173898
Qy 2564 TGTGAATTAATACCTCCAGGCTCACTTTACTGCTGCTGCTGCTGCTGCTGCTGCTG 2623
    |||
Db 173899 TTTGAATTAATACCTCCAGGCTCTTCTGATTTATGCTGCTGCTGCTGCTGCTGCTG 173958
Qy 2624 GGATGAGAGAGCTGCTCAACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2680
    |||
Db 173959 TGTGAATGAGCTGCTCAACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174018
Qy 2681 ACMAAATTTCTGCTAATGACTCCACACTGAGCTTCTTACCTGAACTGATTTATCA 2740
    |||
Db 174019 ACMAACAGAG-TGTTGAAGATCTGACCTAGCTCTCTGCTGCTGCTGCTGCTGCTG 174077
Qy 2741 CCGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2800
    |||
Db 174078 CCGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174137
Qy 2801 CATATCCACATTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2860
    |||
Db 174138 TATATCCAGGCAAAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174197
Qy 2861 ACTGGAATGAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2919
    |||
Db 174198 GATGGAATGAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174257
Qy 2920 CTAATTAATTTCTTAATAAATTTTAACTACAGCAAGTATTTTACTGCTGCTGCTG 2979
    |||
Db 174258 GCTTAATTTCTTAATAAATTTTAACTACAGCAAGTATTTTACTGCTGCTGCTG 174317
Qy 2980 TTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3038
    |||
Db 174318 GGTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174377
Qy 3039 CATTTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3098
    |||
Db 174378 AATTTCTAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174437
Qy 3099 AGATTCTAAATGAACTATCTTTTCAATTCATCTGATCTGCTGCTGCTGCTGCTG 3158
    |||
Db 174438 GATTTTCTAA- - - - -CTTTTAAATAATTTCTGCTGCTGCTGCTGCTGCTG 174488
Qy 3159 AAGAACTGTTAATAGCAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3218
    |||
Db 174489 AAGAACTGTTAACA-CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174547
Qy 3219 TCTTTCATAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3277
    |||
Db 174548 TCTTTCATAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174606
Qy 3278 CATGCAATTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3337
    |||
Db 174607 CATGCAATTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174662
Qy 3338 CCTAATTTGTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3379
    |||
Db 174663 CCTAATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174704

```

RESULT 10  
 AC117867/c AC117867 242679 bp DNA linear HTG 11-OCT-2002  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-37619, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 11 unordered pieces.  
 AC117867  
 HTG, HTGS PHASE1, HTGS DRAFT, HTGS ENRICHED.  
 KEYWORDS  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 242679)  
 Muny, D. Marie, Metzger, M. Lee, Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Albrooke, S., Amin, A., Angiano, D.,  
 Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Bialato, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Burch, P., Burrell, K., Caldwell, R.,  
 Cadenas, V., Carter, K., Cavazo, I., Caesar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyte, M., Cree, A., D'Souza, L.,  
 Davila, M., Davis, C., Day-Carroll, L., De Andrade, C., Dedrich, D.,  
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dimh, R., Divya, K.,  
 Diegel, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Davies, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falis, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,  
 Hollins, B., Howles, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kovis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Loshenshwa, L., Louie, H., Lozada, R., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapa, P., McNeill, T.Z., Meenen, E.,  
 Manthey, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Muniasoon, A., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwokweli, O., Okunnu, G., Olarnunsgoon, A., Pal, S., Parks, K.,  
 Pasternek, S., Paul, H., Perez, A., Peters, L., Pfankuch, C.,  
 Plopper, F., Polindexter, A., Popovic, D., Pritts, E., Pu, L.,  
 Puzo, M., Quatro, J., Rabin, E., Reeves, K., Regier, M.A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., S.J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smale, J.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Swatek, A., Tabot, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willison, R., Wiczyski, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yan, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.  
 Direct Submission  
 2 (bases 1 to 242679)  
 Morley, K.C.  
 Direct Submission  
 Submitted (11-APR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 242679)  
 Rat Genome Sequencing Consortium.  
 Rat Genome Submission  
 Submitted (11-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Oct 9, 2002 this sequence version replaced gi:21746224. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

## ----- Genome Center

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: CT2A  
 Center clone name: CH230-37619

## ----- Summary Statistics

Assembly program: Phrap; version 0.990129  
 Consensus quality: 188097 bases at least Q40  
 Consensus quality: 190770 bases at least Q30  
 Consensus quality: 192614 bases at least Q20  
 Estimated insert size: 191086; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 11 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 185634: contig of 185634 bp in length  
 \* 185635 185734: gap of unknown length  
 \* 185735 189337: contig of 3603 bp in length  
 \* 189338 189437: gap of unknown length  
 \* 189438 198672: contig of 9235 bp in length  
 \* 198673 198772: gap of unknown length  
 \* 198773 202047: contig of 3275 bp in length  
 \* 202048 202147: gap of unknown length  
 \* 202148 203477: contig of 1330 bp in length  
 \* 203478 203577: gap of unknown length  
 \* 203578 204617: contig of 1040 bp in length  
 \* 204618 204717: gap of unknown length  
 \* 204718 205848: contig of 1131 bp in length  
 \* 205849 205948: gap of unknown length  
 \* 205949 207199: contig of 1251 bp in length  
 \* 207200 207299: gap of unknown length  
 \* 207300 208535: contig of 1236 bp in length  
 \* 208536 211249: gap of unknown length  
 \* 211250 211349: contig of 2614 bp in length  
 \* 211350 242679: gap of unknown length  
 \* Location/Qualifiers

## FEATURES

Source  
 1..242679  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-37619"  
 misc\_feature  
 1..1668  
 /note="wgs and extension  
 clone\_end:5p6"  
 1914..2738  
 /note="clone\_boundary  
 clone\_end:5p6  
 site:Mbol

misc\_feature end\_sequence:RXAMW537Y  
 8953..9899  
 /note="clone\_boundary  
 clone\_end:T7  
 site:Mbol  
 misc\_feature end\_sequence:RXAMW537Y  
 185735..187055  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 misc\_feature 188311..189337  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 misc\_feature 198773..199845  
 /note="wgs\_end\_extension  
 clone\_end:T7"  
 BASE COUNT 59286 a 39026 c 39489 g 55897 t 48981 others  
 ORIGIN  
 Query Match 23.6%; Score 819.4; DB 2; Length 242679,  
 Best Local Similarity 83.8%; Pred. No. 1,5e-166;  
 Matches 995; Conservative 0; Mismatches 166; Indels 26; Gaps 5;  
 QY 187 ATACGACATCTGATGATACCTGGGGAATTCGCTCTGACATGATGAAAGT 246  
 Db 239544 AGACAGACCTCTGGGTACCTGGGCAATTTGGCTTTCCTACCGAATTAATGAAAGC 239485  
 QY 247 TGGATATAGACGAGTTGATTT--ATTATGAAGTAGACGCTCACTACCACT 303  
 Db 239484 TGAATGTAGACAGCTGATTTATTAATTCGTGAAGTAGAGCTCACTACAGCGCTT 239425  
 QY 304 CCAGGATTAACACTTTTTCAGCATCACTCACCTGAGCTCTTATCAATTTTGAAT 363  
 Db 239424 TAAACTTCCTGCA--GTGTCAGCAATTTCTTCACTGTGAGCTCTTAAATTTGCTT 239366  
 QY 364 TCTTGGG-----GGAAATATACGGGATAGAGAGGCTATTTTAA 406  
 Db 239365 TATGGGAAAAATAAAAAACAATAAACAACCTGGGTAGAGAGGTAATTAACAAA 239306  
 QY 407 --TAAGTAGCATCTTTTCCCTTTCAAGTTGATCCAAAGAGCTGGAGCTC 464  
 Db 239305 AGAAAGTAGTACCTTTTCTTCACTTACGATGATGAAAGCCAGGCTGTAAACC 239246  
 QY 465 CATTGATTCACCTTTAAATCAAAATAGACAGACAGAAAGGACATGCTCTGA 524  
 Db 239245 AGTTGATTCACCTTTAAAGTTCA--GGAAGCTGCAAGAAAGGACATGCTCTGA 239189  
 QY 525 GTGGAACCTGATGCTTATTTATCTCGAAGAACAAAGGTCCGAGTTCCCACTCTTC 584  
 Db 239188 GTGGAACTGTAGGCGTTATTTATCTCGGAGCAAGAGGCTGCTGCCAATCTTTCC 239129  
 QY 585 CTGAGGTGTAGAGCTGAATGTGGGGGTCAAGTTTATTTATCTCGCAATTCACATTGA 644  
 Db 239128 CTGAAGTCACTAGAGCTGAATGTTGGGGGCAAGTTTATCTTACCTCGCAATTCACATPAA 239069  
 QY 645 TAAGCATCCCTATTCCTCCTGTGAAAAATTTTCCCAAGAGACAGAGCTTAATG 704  
 Db 239068 TAAGTATCCCATTCCTCCTGTGAAAAATTTTCCCAAGAGACAGCTTAACG 239009  
 QY 705 ACTTGAACAAGACCTCAAGGAAAGTTTTTATGAGAGATGATCTTGTTCGGT 764  
 Db 239008 ACTTGAACAAGACCTCAAGGAAAGTTTTTATGAGAGATGATCTTGTTCGGT 238949  
 QY 765 AATATCTGACATCTCAAGAGACAGAGAGTCTCTCGAATCACTTCCAGAAAAAG 824  
 Db 238948 AATATCTGACATCTCAAGAGACAGAGAGTCTCTCGAATCACTTCCAGAAAAAG 238889  
 QY 825 GAAGACTGAAAAAGAAAGCTGAATCTTCAAGCTCCAGACTTGCTCAATCTCGAGCC 884  
 Db 238888 GAAGGCTGAAAAAGAAAGCTGAATCTTCAAGCTCCAGACTTGCTCAATCTCGAGCC 238829  
 QY 885 CCGATGAATCAAGCAAGCCCAAGATGAATTCGCACAGTACTTGAAGATGCTCTCC 944  
 Db 238828 CTGAGGAAGTCAAGCAAGTCCGAGATGAATTCGCACAGTACTTGAAGATGCTCTCC 238769



/db xref="taxon:10090"  
 /clone="RP23-1113"  
 BASE COUNT 69218 A 47164 C 45440 G 68306 T

Query Match 23.4%, Score 811.8; DB 10; Length 230128;  
 Best Local Similarity 83.4%; Pred. No. 6.8e-155;  
 Matches 982; Conservative .0; Mismatches 172; Indels 27; Gaps 4;

Db 1005 GATTTCATTACGTCGGGTTACAGAGATCTGACCTTGCGGCAAGAGGCAAGCAGATG 1064  
 238708 GTTTTATTAACGTCGGGTTACAGAGATCTGACCTTGCGGCAAGAGGCAAGCAGATG 238649  
 Qy 1065 CCAGAGTTTCGAGAGATTCCTCCGAGATTTTGTTGTGAGAGATTTCTTGCCAAAGAG 1124  
 Db 238648 CCAGAGTTTCGAGAGATTCCTCCGAGATTTTGTTGTGAGAGATTTCTTGCCAAAGAG 238589  
 Qy 1125 TCTTTGAGAGAACTTTGAATGAAAGCAGAGACCTTGATTCAGCCCGAGAAAGATACCT 1184  
 Db 238588 TTTTGGTGAACCTTTGAATGAAAGAGAGACCCGAGCCAGCTCAGAAAGATACCT 238529  
 Qy 1185 CCAGATTTTATCTCAATTTCAAGACCTGGAAGAGGCGTTTGTATGTTGTGAGAGTGTG 1244  
 Db 238528 CCAGATTTTATCTCAATTTCAAGACCTGGAAGAGGCGTTTGTATGTTGTGAGAGTGTG 238469  
 Qy 1245 GATTTCACATGTCGCTGTATATCATGTCGTCAGCATCTTTTATCATCAACCAATTAAG 1304  
 Db 238468 GATTTCACATGTCGCTGTATATCATGTCGTCAGCATCTTTTATCATCAACCAATTAAG 238409  
 Qy 1305 AAGACAAATCTGTCACAGTACATGAAATATGTTCTTCAAGTGTG 1351  
 Db 238408 AAGACAAATCTGTCACAGTACATGAAATATGTTCTTCAAGTGTG 238362

RESULT 11  
 AC098707 230128 bp DNA linear ROD 21-JUN-2002  
 LOCUS Mus musculus clone RP23-1113, complete sequence.  
 AC098707 AC098707.2 GI:19909459  
 VERSION HTG.  
 KEYWORDS Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 SOURCE Mus musculus  
 Tissue Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 230128)  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 230128)  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-OCT-2001) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 3 (bases 1 to 230128)  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-APR-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 230128)  
 AUTHORS McPherson,J.D. and Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 COMMENT On Apr 3, 2002 this sequence version replaced gi:16554404.

FEATURES  
 source Location/Qualifiers  
 1..230128  
 /organism="Mus musculus"





clone\_end:T7  
 site:ECORI  
 end\_sequence:BH365326"  
 138060..136295  
 /note="clone\_boundary  
 clone\_end:Sp6  
 site:ECORI  
 end\_sequence:BH365327"

BASE COUNT 69486 a 47001 c 47902 g 74288 t 11026 others  
 ORIGIN

Query Match 23.2% Score 806.2; DB 2; Length 249703;  
 Best Local Similarity 69.2%; Pred.No.1.1e-163;  
 Matches 1420; Conservative 0; Mismatches 573; Indels 60; Gaps 21;

QY 1347 GTGAGCCTCCAGATGGTCACTGACCTGCGATTCGCTGCGAAGATGGCAAGGTG 1406  
 159149 GTGAGCCTCCAGATGGTCTCTCCATTTGTGATTCCTGCTGCAAGATGGCAAGGTG 159208  
 QY 1407 ACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1466  
 159209 ACA--AAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159265  
 QY 1467 AGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1526  
 159266 AGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159325  
 QY 1527 ACATCCAGACTTGACCGTCCCATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1586  
 159326 ACATCCAGACTTGACCGGCGCCATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159385  
 QY 1587 AGATGCGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1446  
 159386 AGATGAGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159445  
 QY 1647 TGAGCAGCAAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1706  
 159446 TAAAGAGCAAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159505  
 QY 1707 GATTCAGAGATTCCTCAAAAAAATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1766  
 159506 GATTCAGAGATTCCTCAAAAAAATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159565  
 QY 1767 GCGCATCTGAATTTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1826  
 159566 GCGAGCTGAATTTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159620  
 QY 1827 AAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1886  
 159621 -----CCACTTTGAATTAACCTCATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159669  
 QY 1887 AATCAACTAATGATGACATTTCTTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1946  
 159670 TA-ACAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 159727  
 QY 1947 CTTTACCTAGTCACTTAACATGTAATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2006  
 159728 CTTTACCTAGTCACTTAACATGTAATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159787  
 QY 2007 GPACAGAGAAATCTTTTAAATTAAGTATTTG-----TGTGTTACTTGTGCGGAGGAGG 2061  
 159788 CCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159847  
 QY 2062 TATCTT-ATAATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2120  
 159848 AATCTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 159907  
 QY 2121 TTAATCCAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2180  
 159908 TGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2240  
 QY 2181 GTAGAAAAAGATCTGCGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2240

DB 159667 GTTGAAGAGATTAACCCAGTGGCATATGTTAAGAAATTAATCTTTTCTTACTTTT 160026  
 QY 2241 TGTTTATGGGTTGGGGGAGATGGCAATTTATATGACTTTTCACTCAATCTATATGTG 2300  
 DB 160027 GTTTATGGGTTGGGGGAGATGGCAATTTATATGACTTTTCACTCAATCTATATGTG 160086  
 QY 2301 CCAGTTTATATGACTCCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2359  
 DB 160087 CCAGTTTATATGACTCCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 160146  
 QY 2360 GTATATACATATGACGACATGACATGACATGACATGACATGACATGACATGACATGAC 2419  
 DB 160147 ATATATACATGAAATGACATGACATGACATGACATGACATGACATGACATGACAT 160206  
 QY 2420 TCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2479  
 DB 160207 TCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 160265  
 QY 2480 TCTTTTGGAGAGATTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2538  
 DB 160266 CCTCTGGAGAGATTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 160325  
 QY 2539 ATCATCAAAATTTGATTCCTTATGTAATTAATTAATTAATTAATTAATTAATTAATTA 2598  
 DB 160326 ATCATCAAAATTTGATTCCTTATGTAATTAATTAATTAATTAATTAATTAATTA 160381  
 QY 2599 TTTGCTCTTTGCTGCTATTAAGAGAGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2658  
 DB 160382 TGTGCTCTTTGCTGCTATTAAGAGAGATGAGAGAGGAGGAGGAGGAGGAGGAGGAG 160441  
 QY 2659 AA--AAATCAAAATTTGATTCCTTATGTAATTAATTAATTAATTAATTAATTAATTA 2715  
 DB 160442 AACAGACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 160500  
 QY 2716 TTTCTTACCTGAGAGGATTTATCACTTTTCCATGTTTTCAGAGTTTCTACTGCCC 2775  
 DB 160501 TTTCTTACCTGAGAGGATTTATCACTTTTCCATGTTTTCAGAGTTTCTACTGCCC 160560  
 QY 2776 ACAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2835  
 DB 160561 ACATTTTATTAAGAGGCTTT-ATATTAATCAAGGCAATTTCTTGATATCTGTG 160619  
 QY 2836 TGTGGAATTAAGAGCAACACA-----GATCTGAATTAAGGAGGAGGAGGAGGAGGAG 2890  
 DB 160620 TGTGGAATTAAGAGCAACACA-----GATCTGAATTAAGGAGGAGGAGGAGGAG 160679  
 QY 2891 TTTAGAGAAATGATTTTACACACAACTTAATTAATTTCTTATTAATTAATTTTAACTAC 2950  
 DB 160680 TTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 160739  
 QY 2951 AAAGCTAATTTTACTTTCCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3009  
 DB 160740 AACTATACCTGTTGTTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 160799  
 QY 3010 GCTGTGCCCATGCTAGATTTAGCTGTGATGATGATGATGATGATGATGATGATGATGAT 3069  
 DB 160800 ACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 160859  
 QY 3070 GGTATCTGAAGCTCCAGAGTTAAGGTTGAGATTTGCAATGAACTATCTTTTCAAT 3129  
 DB 160860 GGTATCTGAAGCTCCAGAGTTAAGGTTGAGATTTGCAATGAACTATCTTTTCAAT 160911  
 QY 3130 ACATCTGATCTGTTATGACACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3189  
 DB 160912 TAGCTGATCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 160970  
 QY 3190 ACTGTGATTTTATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3249  
 DB 160971 ACTGTGATTTTATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 161030  
 QY 3250 AAAGCAAGTTTT-ATTTTCAAGATTCCTCATGATTTTCAAGTGTACCAAAATTAAT 3308  
 DB 161031 AAAGCAAGTTTTTATTTTCAAGATTCCTCATGATTTTCAAGTGTACCAAAAGAG 161090

```

Qy 3309 TGCATTATTAATGTTGTCGCAAGACCTCTTAATTTGTTTATGCGTGTGTCATG 3368
Db 161091 TTT---ATCACTAGTGTGTCGCAAGACCTCTT-ATTTTGTGTTTGTGTGTGTGTG 161146
Qy 3369 TGTGTATGTGTAT 3381
Db 161147 TGTGTGTGTGTGT 161159

RESULT 13
BC049734 781 bp mRNA linear ROD 01-APR-2003
DEFINITION Mus musculus, clone IMAGE:6771233, mRNA.
ACCESSION BC049734
VERSION BC049734.1 GI:29436685
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 781)
AUTHORS Strausberg, R.
JOURNAL Direct Submission
Submitted (31-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NHL-MGC Project URL: http://mgc.nci.nih.gov
REMARK Contact: MGC help desk
Email: gcgdb-remail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
CDNA Library Preparation: Michael Brownstein Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Boedet, Yaron Butcherfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzyminski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candace McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Diane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Natasha van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Zetter.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: lnl.plate: 46 Row: 9 Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
location/Qualifiers
source 1..781
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6771233"
/tissue="Testicle, mouse"
/clone_lib="NHL MGC_169"
/lab_host="DH10B"
/notes="Vector: pDNR-LIB"
BASE COUNT 348 a 146 c 169 g 118 t
ORIGIN
Query Match 10.6%; Score 367.4; DB 10; Length 781;
Best Local Similarity 80.2%; Pred. No. 8.5e-69;
Matches 463; Conservative 0; Mismatches 96; Indels 16; Gaps 2;
Qy 1347 GTGAGCTTCCAGATGTCACCTCTGACCTGATTCCTGTCGAAAGGGAAGG 1406
Db 101 GTGAGCTTCCGAGTGTCTCTCTATGTGACTGCTGCTGCAAGATGGCAAGGAG 160

Qy 1407 ACAAGAGAGGAGAGGAGGACGCTCTTGATGATGACCTCTCCATCTAGTCGACAGC 1466
Db 161 ACN---MAGAGAGAGGAGGACCTCTGCAATGACCTCTCCATCTAGTCGACAGC 217
Qy 1467 AGCTGAGGCGGACGCTCTCCGAGAGACGTCATGTGTGTCGCGGAGACAGCGAGACA 1526
Db 218 AGTCAGAGCGGACGCTCTCCGAGAGACGTCATGTGTGTCGCGGAGACAGCGAGACA 277
Qy 1527 ACATCGACATCTTGACCGCTCCATCAAGAGGCGCTGTCACTGATCAACAGCTCG 1586
Db 278 ACATCGACATCTTGACCGCTCCATCAAGAGGCGCTGTCACTGATCAACAGCTCG 337
Qy 1587 AGATCGGCGGAGAGAGGACCTACTCCGATTTGACTTCAAGGCTCCAGGAAATGACA 1646
Db 338 AGATCGGCGGAGAGAGGACCTACTCCGATTTGACTTCAAGGCTCCAGGAAATGACA 397
Qy 1647 TGAGCAGCAAAAAAAGCTTTAAGAAAGAAAGCTCTCAATTGAGAGAGAGCTGAGAAAT 1706
Db 398 TAAAGCAGCAAAAAAAGCTTTAAGAAAGAAAGCTCTCAATTGAGAGAGAGCTGAGAAAT 457
Qy 1707 GTATCCAGATTTCTTAAGAAAGAAAGCTCTCAATTGAGAGAGAGCTGAGAAAT 1766
Db 458 GTATCCAGATTTCTTAAGAAAGAAAGCTCTCAATTGAGAGAGAGCTGAGAAAT 517
Qy 1767 GGCATCTGAATCTTTAAGAGATATCATCTTAAGAGAGGCGTGGGCGGAGAAAAA 1826
Db 518 GGCATCTGAATCTTTAAGAGATATCATCTTAAGAGAGGCGTGGGCGGAGAAAAA 566
Qy 1827 AAAAAAAGATCTTTAAGAAATTAACCTCAATTAAGAAAGAAATTAATTAAGAA 1886
Db 567 ---GTAATCGCACTTTAAGAAATTAACCTCAATTAAGAAAGAAATTAATTAAGAA 622
Qy 1887 AATPACCTAATGATGACATTTCTTAAGAAACAAATA 1923
Db 623 AAAAAAAGATCTTTAAGAAATTAACCTCAATTAAGAAAGAAATTAATTAAGAA 659

RESULT 14
EX323465 175059 bp DNA linear HTG 06-JUN-2003
LOCUS Dantio rerio clone DKEVP-9C6, *** SEQUENCING IN PROGRESS ***, 3
unrounded pieces.
EX323465
ACCESSION EX323465.4 GI:31559295
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULFILL.
KEYWORDS Dantio rerio (zebrafish)
SOURCE Dantio rerio
ORGANISM Dantio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Osteichthyes;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 175059)
REFERENCE
AUTHORS McLaren, S.
JOURNAL Direct Submission
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Jun 9, 2003 this sequence version replaced gi:29825526.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zkp9c6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 174655 bases at least Q40
Consensus quality: 174734 bases at least Q30
Consensus quality: 174818 bases at least Q20
Insert size: 176859; sum-of-contris
Insert size: 176842; 1.7% error; agarose-fp
Quality coverage: 10.25x in Q20 bases; sum-of-contris Quality

```

coverage: 10.15x in Q20 bases; agarose-fp

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 44665: contig of 44665 bp in length  
 \* 44666 44765: gap of 100 bp  
 \* 44766 172724: contig of 127959 bp in length  
 \* 172725 172824: gap of 100 bp  
 \* 172825 175059: contig of 2235 bp in length.  
 Location/Qualifiers  
 \* 1.175059  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKXP-9C6"  
 /clone\_11b="Daniokeypilot"

## FEATURES

source

misc\_feature  
 1..44665  
 /note="assembly\_fragment:01355  
 clone\_end:17"

misc\_feature  
 44766..172724  
 /note="assembly\_fragment:00935.0"

misc\_feature  
 172825..175059  
 /note="assembly\_fragment:01078"

BASE COUNT 56742 a 30793 c 31197 g 56126 t 201 others  
 ORIGIN

Query Match 9.2%; Score 319.2; DB 2; Length 175059;  
 Best Local Similarity 63.6%; Pred. No. 2.3e-58;  
 Matches 533; Conservative 0; Mismatches 278; Indels 27; Gaps 2;

QY 514 AATGGCTCTAGTGAAGAACTAGTCGTATATTCCTCGAGAAACAAGAGTCCGACGTTC 573  
 Db 127701 AATGGCTCTAGTGAAGAACTAGTCGTATATTCCTCGAGAAACAAGAGTCCGACGTTC 127760  
 QY 574 CAATCTCTCTGAGTGTGACCTGAATGTGGGGGCAAGTTTATTTATCTCCGCA 633  
 Db 127761 GAGTGTCTCTGAGTGTGACCTGAATGTGGGGGCAAGTTTATTTATCTCCGCA 127820  
 QY 634 TTCCACATTGATAGCATCTCTCATTCCTCTGTGAAAAATGTTTCCCAAGAGCA 693  
 Db 127821 TTCCACATTGATAGCATCTCTCATTCCTCTGTGAAAAATGTTTCCCAAGAGCA 127880  
 QY 694 CACGGCTAATGATCTAGCCAGAGCTCCAGAGGATTTTTCATTGACAGATGAT 753  
 Db 127881 CATTCTTAAGACCTCAGCCAGAGATCAGAGGACGCTACTTATGACCGGAGCGAT 127940  
 QY 754 CTGTTCGTTATATTCGACATCTCAGGAGCAGGAGTGTCTGCTGATCACTT 813  
 Db 127941 TCTCTTAAAGTACGTCTGACATCTCAGGAGTACAGTGTCTGCTGCGGATTTT 128000  
 QY 814 TCCGAAAAAGAAAGCTGAAAAAGGAGAGTAACTTCCACCTCCGAGCTGGTCAA 873  
 Db 128001 TCCGAAAAAGGAGAGCTGAAAAAGGAGAGTAACTTTCACCTCCGAGCTGGTCAA 128060  
 QY 874 ACTCTGACCCCGAGTAATCAAGCAAGCCCGAGTAACTTTCGACAGAGTATTTGA 933  
 Db 128061 AATCTCT-----AAACCCAGATGATTAATGACAGAGATTTGA 128099  
 QY 934 AGATGCTCTCCAGAGAGCAGACAGAAATCTGCCCTTCTCTCTGCTGCTGCGCA 993  
 Db 128100 CGAGAGATCCAGAGGAGAGCAGACAGAGTATATTCAGCTCTTACCTGAGCGCCGGA 128159  
 QY 994 CGGCAAGTGGGTTTCACTTATCTGTGGTTTACAGAGATCTCTGACCTTGGGAGAGAGG 1053  
 Db 128160 CCGAGCTACGGCTTCACTGCTGTGGGTTACAGAGATCTCTGTGCGATTCGGAGAGGACA- 128218

QY 1054 ACAGGAGATGCCAAGTTTCGAGAGTTCCTCCGATTTTGTGTGAGAGATTTCCTT 1113  
 Db 128219 -----CTGATTCCTCAAGAGCCCGGATATACCAAAATCTTCATTCGGAAGAGCGTCT 128273  
 QY 1114 GCGAAAGAGATCTTGTGAGAACTTTGATTAAGACAGACCTGTATGAGCCGCA 1173  
 Db 128274 GCGAAAGAGATCTTGTGAGAACTTTGATTAAGACAGACCTGTATGAGCCGCA 128333  
 QY 1174 AAGTATACCTCCAGATTTTATCTCAATTAAGCACTCTGAGAAAGGCTTTTGTATGTT 1233  
 Db 128334 GCGTTACACTTCTCAGTTTATCTGAAAGTTTCCACCTGAGAGCGGTTGATGCT 128393  
 QY 1234 GTCAAGTGTGATTCACATGTTGGCTGTAACTGATGATGATGATGATGATGATGAT 1293  
 Db 128394 CCGGAGAGAGGAGGTTTCCATGCTGCGGTCAATTCATCACCACATCTCTCAACA 128453  
 QY 1294 CCAATATACAGATGACAGATCTGTCAAGCTACATGATATGTTCTTACCGGTGAG 1351  
 Db 128454 CAGGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 128511

## RESULT 15

BX470157 200467 bp DNA linear HTG 05-MAY-2003  
 LOCUS BX470157  
 DEFINITION Danio rerio, clone CH211-119P14, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 8  
 UNORDERED PIECES.  
 ACCESSION BX470157.2 GI:30387082  
 VERSION HTG; HTGS\_PHASE1.  
 KEYWORDS Danio rerio (zebrafish)  
 SOURCE Danio rerio  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 200467)  
 McLeay, K.

## REFERENCE

## AUTHORS

## JOURNAL

## COMMENT

Direct Submission  
 Submitted (04-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On May 5, 2003 this sequence version replaced gi:30349786.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: zfish-help@sanger.ac.uk  
 Project information  
 Center project name: zc119P14  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 198546 bases at least Q40  
 Consensus quality: 199010 bases at least Q30  
 Consensus quality: 199314 bases at least Q20  
 Insert size: 199767; sum-of-contigs  
 Insert size: 201190; 3.3% error; agarose-fp  
 Quality coverage: 5.95x in Q20 bases; sum-of-contigs Quality  
 coverage: 6.07x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 9514: contig of 9514 bp in length  
 \* 9515 9614: gap of 100 bp  
 \* 9615 14582: contig of 4968 bp in length  
 \* 14583 14682: gap of 100 bp  
 \* 14683 18933: contig of 4251 bp in length  
 \* 18934 19033: gap of 100 bp





```
/ Sequence 1, Application US/09328965
/ Patent No. 6501008
/ GENERAL INFORMATION:
/ APPLICANT: Nevins, Donald J.
/ APPLICANT: Simmons, Carl
/ TITLE OF INVENTION: The Regents of the University of California
/ FILE REFERENCE: 023070-096600US
/ CURRENT APPLICATION NUMBER: US/09/328,965
/ EARLIER FILING DATE: 1999-06-09
/ NUMBER OF SEQ ID NOS: 3
/ SOFTWARE: PatentIn Ver. 2.1
/ LENGTH: 1091
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: maize coleoptile endo-1,3,1,4-beta glucanase cDNA
/ NAME/KEY: CDS
/ LOCATION: (68)..(979)
/ OTHER INFORMATION: endo-1,3,1,4-beta glucanase
US-09-328-965-1

Query Match          2.1%; Score 73.8; DB 4; Length 1091;
Best Local Similarity 65.5%; Pred. No. 3.2e-08;
Matches 108; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 3304 TAAATTGTCATTAAAGTGTGTGCGCAGCACTCCTAAATTTGTTATTCGTGTGCT 3363
DB 918 TAAATTCCTTCATTTTCTTCTCTATGATTTCTTTCTTTCTTTCTTTCTTTTAT 977
QY 3364 GCATGTGTATGTATGATCAGCAAGTAATAAGCAATTGATGATTAATAAAAAA 3423
DB 978 GATCGCAATTAAGTTAGAGCGGTAAAAAATAAAAAAATAAAAAAATAAAAAA 1037
QY 3424 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
DB 1038 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1082

RESULT 3
US-09-620-312D-706/C
Sequence 706, Application US/09620312D
Patent No. 6569662
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Weinman, Tom
/ APPLICANT: Xue, Aiyong J.
/ APPLICANT: Yang, Yonghong
/ APPLICANT: Zhou, Jian-Rui
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yundang
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: John Fillinghaas
/ APPLICANT: Dormanac, Radoje T.
/ TITLE OF INVENTION: No. 6569662el Nucleic Acids and
/ FILE REFERENCE: 784CIP28
/ CURRENT APPLICATION NUMBER: US/09/620,312D
/ EARLIER FILING DATE: 2000-07-19
/ PRIOR APPLICATION NUMBER: 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: 09/488,725
```

```
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 1105
/ SOFTWARE: PL_FL_genes Version 1.0
/ SEQ ID NO 706
/ LENGTH: 4055
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (2515)..(3519)
US-09-620-312D-706

Query Match          2.1%; Score 73.4; DB 4; Length 4055;
Best Local Similarity 58.4%; Pred. No. 6.9e-08;
Matches 128; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 3250 AAAGACCAAGTTTATTTTTCAGATTCCTCAGATTCCTAGTGTAAACAAAAATATTT 3309
DB 500 AAAGCAAAATGTTTTCAGATTTGTTATTAATAAGTTATCCATACCAATAAAAA 441
QY 3310 GTCAATTATAGTTGTGTGCGCAAGCACTCCTAATTTGTTTATTCGCTGTGTCATGT 3369
DB 440 GTGTCAACACAGCATTTTCTGTAAATTAATTAATGTTTCAGTTGTAATTTGTAATTT 381
QY 3370 GTGTAATGTATTCACAGTAATAAGCAATTGATGATTAATAAAAAAATAAAAAA 3429
DB 380 TTTCGCAATCGCTTTATTTATTTATTAATGCTTTTGAATAAAAAAATAAAAAA 321
QY 3430 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
DB 320 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 282

RESULT 4
US-09-996-243-114
Sequence 114, Application US/09996243
Patent No. 6478825
/ GENERAL INFORMATION:
/ APPLICANT: Ashkenazi, Avi J.
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Botstein, David
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gerlitsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Kijavits, Ivar J.
/ APPLICANT: Napier, Mary A.
/ APPLICANT: Pan, James
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ FILE REFERENCE: P2730PIC13
/ CURRENT APPLICATION NUMBER: US/09/996,243
/ EARLIER FILING DATE: 2001-11-14
/ PRIOR APPLICATION NUMBER: 60/049787
/ PRIOR FILING DATE: 1997-06-16
/ PRIOR APPLICATION NUMBER: 60/062250
/ PRIOR FILING DATE: 1997-10-17
/ PRIOR APPLICATION NUMBER: 60/065186
/ PRIOR FILING DATE: 1997-11-12
```

Sun 'Sep 7 09:26:01 2003

us-10-056-884-1.rtf

**Page 3**

PRIOR APPLICATION NUMBER: 60/065311	PRIOR FILING DATE: 1998-06-15
PRIOR APPLICATION NUMBER: 60/066770	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/075945	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/078910	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/083322	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/084600	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/087106	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/087609	PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/087759	PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/087827	PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/088021	PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/088025	PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/088026	PRIOR FILING DATE: 1998-06-20
PRIOR APPLICATION NUMBER: 60/088028	PRIOR FILING DATE: 1998-06-20
PRIOR APPLICATION NUMBER: 60/088029	PRIOR FILING DATE: 1998-06-20
PRIOR APPLICATION NUMBER: 60/088030	PRIOR FILING DATE: 1998-06-20
PRIOR APPLICATION NUMBER: 60/088033	PRIOR FILING DATE: 1998-06-20
PRIOR APPLICATION NUMBER: 60/088326	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088157	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088202	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088212	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088217	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088655	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088734	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088738	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088742	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/088824	PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/088826	PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/088858	PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/088861	PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/088876	PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/089105	PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/089440	PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/089512	PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/089514	PRIOR FILING DATE: 1998-07-01



```

; Sequence 3, Application US/09083587
; Patent No. 6492138
; GENERAL INFORMATION:

```

APPLICANT: Schmandt, et al.  
TITLE OF INVENTION: NOVEL SHC BINDING PROTEIN  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,587  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 01017/34451  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2246 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: /desc = "mouse PAL cDNA"  
US-09-083-587-3

Query Match 2.0%; Score 69.2; DB 4; Length 2246;  
Best Local Similarity 64.2%; Pred. No. 5.4e-07;  
Matches 104; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 3307 TTTGCAATTATAGTGTGTGCGACGACCTCTTAATTTGTTTATGCGTGTGTGCA 3366  
DB 2054 TTGACATGTCTTAAGTATGCTGCTTAATATCTTGTCTTCAATGCGTGTGTA 2113  
QY 3367 TGTGTATGTATGACAGGTAATAAGCAATTGATGATTTAAAAA 3426  
DB 2114 TTATATTAAGTGTACTTGACCAAAAAAAAAAAAAAAAAAAAA 2173  
QY 3427 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468  
2174 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2215

RESULT 8  
US-09-594-506-37  
Sequence 37, Application US/09594506  
Patent No. 6512164  
GENERAL INFORMATION:  
APPLICANT: Famodu, Omojaye O.  
APPLICANT: Rafalski, J. Antoni  
TITLE OF INVENTION: Thiamine Biosynthetic Enzymes  
FILE REFERENCE: BB1372 US NA  
CURRENT APPLICATION NUMBER: US/09/594,506  
CURRENT FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: 60/139,556  
PRIOR FILING DATE: 1999-06-16  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 37  
LENGTH: 2406  
TYPE: DNA  
ORGANISM: Triticum aestivum  
US-09-594-506-37

Query Match 2.0%; Score 69.2; DB 4; Length 2406;  
Best Local Similarity 56.6%; Pred. No. 5.6e-07;  
Matches 128; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 3243 AAGTTAAAGACCGCTTTATTTTCAGCATCTTCACGATTTCACTGATACCAAAA 3302  
DB 2179 AACTCATTAAGCCCAATATTTTGTAGGATCCATTCATCTGCAAGCATTCACA 2238  
QY 3303 ATAAATTTGCAATTAATAGTTGTGTGCGAAGCACTCTTAATTTGTTTATGCGTGTG 3362  
DB 2239 TGTCTGTGTAATTTACTTTTACACCTATCTTGTACATTTCTAATAAGTAAATA 2298  
QY 3363 TGCATGTGTATGTATGATCAAGGTAATAAGCAATTGATGATTAATAAAAAA 3422  
DB 2299 TAAAGATGTAACTAATGATGACAGAAAAAAAAAAAAAAAAAAAAAAAAAGCAAAA 2358  
QY 3423 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468  
DB 2359 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2404

RESULT 9  
US-09-465-558-59  
Sequence 59, Application US/0946558  
Patent No. 6436557  
GENERAL INFORMATION:  
APPLICANT: Morakinyo, Layo O.  
APPLICANT: Orozco Jr, Emil M.  
TITLE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES  
FILE REFERENCE: BB1322 US NA  
CURRENT APPLICATION NUMBER: US/09/465,558  
CURRENT FILING DATE: 1999-12-17  
EARLIER APPLICATION NUMBER: 60/112,734  
PRIOR FILING DATE: 1998-12-18  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 59  
LENGTH: 2202  
TYPE: DNA  
ORGANISM: Glycine max  
US-09-465-558-59

Query Match 2.0%; Score 69; DB 4; Length 2202;  
Best Local Similarity 69.9%; Pred. No. 6e-07;  
Matches 93; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3336 CTCCTAATTTGTTTATTTGCGTGTGTGCAATGTGTATGTATACAGTAATTAAG 3395  
DB 2064 CTCCTGTTTGTTCCTGCTGTGATCTGTAAATTAATACGTGATTTAAG 2123  
QY 3396 GCAATTGATGATTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3455  
DB 2124 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2183  
QY 3456 AAAAAAAAAAAAAA 3468  
DB 2184 AAAAAAAAAAAAAA 2196

RESULT 10  
US-08-665-716-1  
Sequence 1, Application US/08665716  
Patent No. 5789222  
GENERAL INFORMATION:  
APPLICANT: KELLY, ROSEMARIE  
APPLICANT: REGISTER, ELIZABETH A  
APPLICANT: MASUREKAR, PRAKASH S  
TITLE OF INVENTION: P5C REDUCTASE GENE FROM ZALARION  
TITLE OF INVENTION: ARBORICOLA  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MERCK & CO., INC.

STREET: 126 E. LINCOLN AVENUE  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: US  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,716  
FILING DATE: 23-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: KORESEN, ELLIOTT  
REGISTRATION NUMBER: 32,705  
REFERENCE/DOCKET NUMBER: 19453PV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-5493  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1147 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 47..960  
US-08-665-716-1

Query Match  
Best Local Similarity 75.9%; Score 68.8; DB 1; Length 1147;  
Matches 85; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 3357 TGTGTGCATGTCGTATGTCATCAAGCAATGATGATTAATAAAAA 3416  
DB 1021 TCCGTAGACATGTCACAGAGTTCTGGGCTATATAAAAGTTTCATTATGAAAAA 1080

QY 3417 AA 3468  
1081 AA 1132

RESULT 11  
US-09-182-816-22  
Sequence 22, Application US/09182816  
GENERAL INFORMATION:  
APPLICANT: Wisniewski, Nancy  
APPLICANT: Silver, Gary M.  
APPLICANT: Lo, Katherine C.  
TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,  
FILE REFERENCE: FC-3-C1  
CURRENT APPLICATION NUMBER: US/09/182,816  
CURRENT FILING DATE: 1998-10-29  
EARLIER APPLICATION NUMBER: 08/989,510  
EARLIER FILING DATE: 1997-12-12  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 22  
LENGTH: 1736  
TYPE: DNA  
ORGANISM: Ctenocephalides felis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (159) . (1553)

US-09-182-816-22

Query Match  
Best Local Similarity 62.6%; Score 68.6; DB 3; Length 1736;  
Matches 107; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 3298 AAAAAATATTGTCATTAATAGTTGTGTGCGACAGCTCTTAATTTGTTATGGCT 3357  
DB 1565 AATAAATATTGTCATTAATATATATGTTAAATATATGTTACTGTGAATAA 1624

QY 3358 GTGTGTGCATGTCGTATGTCATCAAGCAATGATGATTAATAAAAA 3417  
DB 1625 CGATATGATTTTATTTCAACTGTCAATATTAATAAAAAAAAAAAAA 1684

QY 3418 AA 3468  
DB 1685 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1735

RESULT 12  
US-09-182-816-24/C  
Sequence 24, Application US/09182816  
GENERAL INFORMATION:  
APPLICANT: Wisniewski, Nancy  
APPLICANT: Silver, Gary M.  
APPLICANT: Lo, Katherine C.  
TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,  
FILE REFERENCE: FC-3-C1  
CURRENT APPLICATION NUMBER: US/09/182,816  
CURRENT FILING DATE: 1998-10-29  
EARLIER APPLICATION NUMBER: 08/989,510  
EARLIER FILING DATE: 1997-12-12  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 24  
LENGTH: 1736  
TYPE: DNA  
ORGANISM: Ctenocephalides felis  
US-09-182-816-24

Query Match  
Best Local Similarity 62.6%; Score 68.6; DB 3; Length 1736;  
Matches 107; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 3298 AAAAAATATTGTCATTAATAGTTGTGTGCGACAGCTCTTAATTTGTTATGGCT 3357  
DB 172 AATAAATATTGTCATTAATATATATGTTAAATATATGTTACTGTGAATAA 113

QY 3358 GTGTGTGCATGTCGTATGTCATCAAGCAATGATGATTAATAAAAA 3417  
DB 112 CGATATGATTTTATTTCAACTGTCAATATTAATAAAAAAAAAAAAA 53

QY 3418 AA 3468  
DB 52 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2

RESULT 13  
US-09-471-528-22  
Sequence 22, Application US/09471528  
GENERAL INFORMATION:  
APPLICANT: Wisniewski, Nancy  
APPLICANT: Silver, Gary M.  
APPLICANT: Lo, Katherine C.  
TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
FILE REFERENCE: FC-3-C1-1  
CURRENT APPLICATION NUMBER: US/09/471,528  
CURRENT FILING DATE: 1999-12-27

EARLIER APPLICATION NUMBER: 09/182,816  
EARLIER FILING DATE: 1998-10-29  
EARLIER APPLICATION NUMBER: 08/989,510  
EARLIER FILING DATE: 1997-12-12  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 22  
LENGTH: 1736  
TYPE: DNA  
ORGANISM: Ctenocephalides felis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (159)..(1553)  
US-09-471-528-22

Query Match 2.0%; Score 68.6; DB 3; Length 1736;  
Best Local Similarity 62.6%; Pred. No. 6.8e-07;  
Matches 107; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 3298 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 3357  
DB 1565 AATTAATTTTGTGATTAATATATATATATTAATAATATGTAATTACTGTGAATATAA 1624  
DB 3358 GTGTGTCATGTGTGATGTATGTCACAGTATATAAGCAATTGATGATTAATAAAAA 3417  
DB 1625 CGATATGATTTTATTTTCAAACTGTCAATATATATATATATATATATATATATAT 1684  
QY 3418 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 3468  
DB 1685 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 1735

RESULT 14  
US-09-471-528-24/c  
Sequence 24, Application US/09471528  
Patent No. 6153397  
GENERAL INFORMATION:  
APPLICANT: Wisniewski, Nancy  
APPLICANT: Silver, Gary M.  
APPLICANT: Lo, Katherine C.  
APPLICANT: Brandt, Kevin S.  
TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
FILE REFERENCE: FC-3-CI-1  
CURRENT APPLICATION NUMBER: US/09/471,528  
CURRENT FILING DATE: 1999-12-27  
EARLIER APPLICATION NUMBER: 09/182,816  
EARLIER FILING DATE: 1998-10-29  
EARLIER APPLICATION NUMBER: 08/989,510  
EARLIER FILING DATE: 1997-12-12  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 24  
LENGTH: 1736  
TYPE: DNA  
ORGANISM: Ctenocephalides felis  
US-09-471-528-24

Query Match 2.0%; Score 68.6; DB 3; Length 1736;  
Best Local Similarity 62.6%; Pred. No. 6.8e-07;  
Matches 107; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 3298 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 3357  
DB 172 AATTAATTTTGTGATTAATATATATATATTAATAATATGTAATTACTGTGAATATAA 113  
QY 3358 GTGTGTCATGTGTGATGTATGTCACAGTATATAAGCAATTGATGATTAATAAAAA 3417  
DB 112 CGATATGATTTTATTTTCAAACTGTCAATATATATATATATATATATATATATAT 53  
QY 3418 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 3468  
DB 52 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 1735

RESULT 15  
US-09-634-530-22  
Sequence 22, Application US/09634530  
Patent No. 6290358  
GENERAL INFORMATION:  
APPLICANT: Wisniewski, Nancy  
APPLICANT: Silver, Gary M.  
APPLICANT: Lo, Katherine C.  
APPLICANT: Brandt, Kevin S.  
TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
FILE REFERENCE: FC-3-CI-1  
CURRENT APPLICATION NUMBER: US/09/634,530  
CURRENT FILING DATE: 2000-08-08  
PRIOR APPLICATION NUMBER: 09/471,528  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: 09/182,816  
PRIOR FILING DATE: 1998-10-29  
PRIOR APPLICATION NUMBER: 08/989,510  
PRIOR FILING DATE: 1997-12-12  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 22  
LENGTH: 1736  
TYPE: DNA  
ORGANISM: Ctenocephalides felis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (159)..(1553)  
US-09-634-530-22

Query Match 2.0%; Score 68.6; DB 3; Length 1736;  
Best Local Similarity 62.6%; Pred. No. 6.8e-07;  
Matches 107; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 3298 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 3357  
DB 1565 AATTAATTTTGTGATTAATATATATATATTAATAATATGTAATTACTGTGAATATAA 1624  
QY 3358 GTGTGTCATGTGTGATGTATGTCACAGTATATAAGCAATTGATGATTAATAAAAA 3417  
DB 1625 CGATATGATTTTATTTTCAAACTGTCAATATATATATATATATATATATATATAT 1684  
QY 3418 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 3468  
DB 1685 AAAAAATATTGTCATTAATAGTGTGCGCAGACACTCTTAATTGTTTATTCGCGT 1735

Search completed: September 6, 2003, 08:48:45  
Job time: 227 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2003, 06:58:53 ; Search time 836 Seconds  
(without alignments)  
9541.055 Million cell updates/sec

Title: US-10-056-884-1  
Perfect score: 3468  
Sequence: 1 caagcactgcgcacgaagtcgt.....aaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1537136 seqs, 114988732 residues

Total number of hits satisfying chosen parameters: 3074272

Minimum DB seq length: 0  
Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications NA:\*

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/1/pubpna/PCRT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/1/pubpna/PCRTS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 13: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3468	100.0	3468	US-10-056-884-1	Sequence 1, Appl1
2	769	22.2	769	US-10-056-884-3	Sequence 3, Appl1
3	205	5.9	632	US-10-060-036-4467	Sequence 4467, Ap
4	201	5.8	614	US-10-060-036-564	Sequence 564, Ap
5	167	4.8	2052	US-10-080-980-1	Sequence 1, Appl1
6	104.6	3.0	688	US-10-080-980-8	Sequence 8, Appl1
7	87	2.5	249	US-09-918-995-2111	Sequence 2111, Ap
8	80	2.3	80	US-10-056-884-6	Sequence 45, Appl1
9	79	2.3	425	US-09-814-975-451	Sequence 112, Ap
10	79	2.3	1492	US-09-925-299-112	Sequence 112, Ap
11	79	2.3	1492	US-09-925-299-112	Sequence 112, Ap
12	79	2.3	1493	US-10-106-699-1525	Sequence 1525, Ap
13	76.6	2.2	361	US-10-198-846-8619	Sequence 8619, Ap
14	75	2.2	442	US-09-770-444-872	Sequence 872, Ap
15	74	2.1	664	US-09-739-254-66	Sequence 66, Appl1
16	74	2.1	664	US-09-304-615-66	Sequence 66, Appl1

17	74	2.1	664	US-10-055-098-66	Sequence 66, Appl1
18	74	2.1	664	US-10-054-888-66	Sequence 66, Appl1
19	73.8	2.1	1992	US-10-036-542-18	Sequence 18, Appl1
20	73.4	2.1	2440	US-10-106-698-2024	Sequence 2024, Ap
21	73.4	2.1	4055	US-10-037-270-706	Sequence 706, Ap
22	73	2.1	465	US-10-198-846-1482	Sequence 1482, Ap
23	73	2.1	1814	US-09-918-995-31149	Sequence 31149, A
24	73	2.1	1814	US-10-163-866-23	Sequence 23, Appl1
25	72.6	2.1	405	US-09-918-995-37095	Sequence 37095, A
26	72.2	2.1	297	US-09-764-846-95	Sequence 95, Appl1
27	72.2	2.1	297	US-10-091-483-95	Sequence 95, Appl1
28	72	2.1	165	US-09-960-352-11420	Sequence 11420, A
29	72	2.1	611	US-10-198-846-8503	Sequence 8503, Ap
30	72	2.1	926	US-09-948-820-13	Sequence 13, Appl1
31	71.6	2.1	265	US-09-764-846-24	Sequence 24, Appl1
32	71.6	2.1	265	US-10-091-483-24	Sequence 24, Appl1
33	71.4	2.1	394	US-10-198-846-3918	Sequence 3918, Ap
34	71.4	2.1	420	US-09-918-995-17811	Sequence 17811, A
35	71.4	2.1	600	US-10-198-846-8434	Sequence 8434, Ap
36	71.4	2.1	4064	US-09-873-737A-3	Sequence 3, Appl1
37	71.2	2.1	368	US-09-814-975-47	Sequence 47, Appl1
38	71.2	2.1	498	US-10-106-698-1594	Sequence 1594, Ap
39	71.2	2.1	1672	US-10-198-846-7225	Sequence 7225, Ap
40	71.2	2.1	1672	US-10-063-735-17	Sequence 17, Appl1
41	71.2	2.1	1672	US-10-063-547-17	Sequence 17, Appl1
42	71.2	2.1	1672	US-10-063-516-17	Sequence 17, Appl1
43	71.2	2.1	1672	US-10-063-502-17	Sequence 17, Appl1
44	71.2	2.1	1672	US-10-063-518-17	Sequence 17, Appl1
45	71.2	2.1	1672	US-10-063-518-17	Sequence 17, Appl1

## ALIGNMENTS

RESULT 1  
US-10-056-884-1  
Sequence 1, Application US/10056884  
Publication No. US20030032786A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL BETA-SUBUNIT  
FILE REFERENCE: D0076 NP  
CURRENT FILING DATE: 2002-01-24  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: US 60/263, 872  
PRIOR FILING DATE: 2001-01-24  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 3468  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (515) .. (1798)  
US-10-056-884-1

Query Match 100.0%; Score 3468; DB 14; Length 3468;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAGCAGCTGCTAAGTCTTTTCATATGCAAGAAAGTTGCGCAAAATTAATGCT 60  
DB 1 CAAGCAGCTGCTAAGTCTTTTCATATGCAAGAAAGTTGCGCAAAATTAATGCT 60  
QY 61 TTGAACATGAGCAATTTTCTCTACCGCTAGCTATATCCCAAGCATCATGAAGTGA 120  
DB 61 TTGAACATGAGCAATTTTCTCTACCGCTAGCTATATCCCAAGCATCATGAAGTGA 120  
QY 121 GATTCGACAGCTGTGTATTTTCAAGTTCACATGAATGAACCTGACATATATGTTAA 180

[illegible][illegible]



```
Db 601 ACCCGAGTGGGTTTCATTCTGCGGTTCACAGAGATCTTCACCTTGGCGAGAGAG 660
Qy 1053 GACAGCGAGATGCGCAAGTTTCGAGAGTTCCCGGATTTTGGTTTGGAGGATTTCTT 1112
Db 661 GACAGCGAGATGCGCAAGTTTCGAGAGTTCCCGGATTTTGGTTTGGAGGATTTCTT 720
Qy 1113 TGGCAAAAGAGTCTTTGGAGAACTTTGAATGAAGACAGACCTTGA 1161
Db 721 TGGCAAAAGAGTCTTTGGAGAACTTTGAATGAAGACAGACCTTGA 769

RESULT 3
US-10-060-036-4467
; Sequence 4467, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060.036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4467
; LENGTH: 632
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 552..569
; OTHER INFORMATION: n = A,T,C or G
US-10-060-036-4467

Query Match
Best Local Similarity 93.4%; Score 205; DB 14; Length 632;
Matches 214; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 3206 AGTACCAATAGCTCTTTCATATGACTTGTGCTACAGAGGTTAAAGACAGCTTTTAT 3265
; 15 AGTACCAATAGCTCTTTCATATGACTTGTGCTACAGAGGTTAAAGACAGCTTTTAT 74
Qy 3266 TTCAGATTCTCTCATGATTTGATGAGTACCAAAATATTTGTCAATTATAGTTGT 3325
Db 75 TTCAGATTCTCTCATGATTTGATGAGTACCAAAATATTTGTCAATTATAGTTGT 134
Qy 3326 GTGCCAAGACTCTCTATTTGTTTATGCGGTGTGTGATGATGTGTATGATCACA 3385
Db 135 GTGCCAAGACTCTCTATTTGTTTATGCGGTGTGTGATGATGTGTATGATCACA 194
Qy 3386 GGTATATTAAGCATTTGATGATTTAAATTAATTAATTAATTAATTAATTAATTA 3434
Db 195 GGTATATTAAGCATTTGATGATTTAAATTAATTAATTAATTAATTAATTAATTA 243

RESULT 4
US-10-060-036-564
; Sequence 564, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
US-10-060-036-564

TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060.036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 564
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 534..551, 575, 576
; OTHER INFORMATION: n = A,T,C or G
US-10-060-036-564

Query Match
Best Local Similarity 93.3%; Score 201; DB 14; Length 614;
Matches 210; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 3210 ACCAATAGCTCTTTCATATGACTTGTGCTACAGAGGTTAAAGACAGCTTTTATTTCA 3269
Db 1 ACCAATAGCTCTTTCATATGACTTGTGCTACAGAGGTTAAAGACAGCTTTTATTTCA 60
Qy 3270 GCATTCCTCATGATTTGATGAGTACCAAAATATTTGTCAATTATAGTTGTGTC 3329
Db 61 GCATTCCTCATGATTTGATGAGTACCAAAATATTTGTCAATTATAGTTGTGTC 120
Qy 3330 CAGACCTCTCATGATTTGATGAGTACCAAAATATTTGTCAATTATAGTTGTGTC 3389
Db 121 CAGACCTCTCATGATTTGATGAGTACCAAAATATTTGTCAATTATAGTTGTGTC 180
Qy 3390 ATTAAGCATTTGATGATTTAAATTAATTAATTAATTAATTAATTAATTAATTA 3434
Db 181 ATTAAGCATTTGATGATTTAAATTAATTAATTAATTAATTAATTAATTAATTA 225

RESULT 5
US-10-080-980-1
; Sequence 1, Application US/10080980
; Publication No. US20030036115A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN PORRASSIUM CHANNEL BETA-SUB
; TITLE OF INVENTION: K-betaLAME, EXPRESSED HIGHLY IN THE SMALL INTESTINE
; FILE REFERENCE: D0121 NP
; CURRENT APPLICATION NUMBER: US/10/080.980
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/270.132
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/278.953
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (121)..(1095)
US-10-080-980-1

Query Match
Best Local Similarity 4.8%; Score 167; DB 14; Length 2052;
Matches 267; Conservative 0; Mismatches 140; Indels 6; Gaps 1;

Qy 967 CCCCCCTTCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1026
Db 705 CAGCGCGTCCAGTGGTGGAGCGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 764
Qy 1027 AGGATCTGCACCTTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1086
```



Db 765 CGGCTCTACACCACTGGGCGGAGCGCGAGCGGAGCCAGATCTCCGGCGAGTGCGCG 824  
Qy 1087 GAATTTGGTTTGTGAAGAATTTCTTGTGCAAAAGAAAGCTTTTGAAGAACTTTGAATGA 1146  
Db 825 CATCACCGTTTGGCGAAGACGTGCTGCGCAAGAGAGTGTGGGAGAACCTCTAAGGA 884  
Qy 1147 AAGCAGAGACCTGTAGAGCCCGCAGAAAGATACACCTCCAGATTTTATCTCAATTCGA 1206  
Db 885 AAGCCGGAGACCCGACCGTCCCGGAGCGCTACACCTCGCGCTATTAATCAAGTTCAA 944  
Qy 1207 GCACCTGGAAGGCTTTTATATGTGTGTCAGAGTGTGAATTCACATGTGTGCTGTA 1266  
Db 945 CTTCCTGAGAGAGGCTTTCAGAACGCTGTCCGAGTCCGAGCTTCCATGTGTGCTGAG 1004  
Qy 1267 CTGATCGGTGACAGATCTTT-----CATCAACCATATATACAGATGACAAATCTGATC 1320  
Db 1005 CTCACGGGACCTGCGCGCTTTTCCAGAGCAACGACAGAGGAGGAGACAAATCTGAGAC 1064  
Qy 1321 AAGCTACACTGAATATATGTCTTCTTACCGTGAAGCTTCCAGATGTCAACCTTCAC 1373  
Db 1065 CAGCTACACCGAGTACGTCTTGTGAGGAGTGAAGCTCCCGACGCCCTCGC 1117

SUBJ 6  
US-10-080-980-8  
Sequence 8, Application US/10080980  
Publication No. US20030036115A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL BETA-SUB  
FILE REFERENCE: D0121 NP  
CURRENT FILING DATE: 2002-02-21  
PRIORITY APPLICATION NUMBER: US/10/080,980  
PRIORITY APPLICATION NUMBER: US 60/270,132  
PRIORITY FILING DATE: 2001-02-21  
PRIORITY APPLICATION NUMBER: US 60/278,953  
PRIORITY FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: Patent in version 3.0  
SEQ ID NO 8  
LENGTH: 688  
TYPE: DNA  
ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: wherein "N" is equal to "A", "C", "G" or "T".  
US-10-080-980-8

Query Match 3.0%; Score 104.6; DB 14; Length 688;  
Best Local Similarity 51.1%; Pred. No. 1,7e-14;  
Matches 192; Conservative 0; Mismatches 178; Indels 6; Gaps 1;

Qy 563 TCCGCAATTCCCAATCTCTTCTGAGGTGTAAGCTGAATGTGGGGGTCAAGTTAT 622  
Db 186 TCCGCGAGAGCCACGCGCTTCTCCGACATCGTGAAGCTGAACGTGGGGGCCAGAGGTATC 245  
Qy 623 TTTACTCGCAATTCACATGATTAAGCATCCCTCATTTCTCTGTGAAAAATGTTTTCC 682  
Db 246 GTGACCCCGCGCTGACAGGTGTGTGCGCCGACGTGCTGCTCGGCGCAGTTCACG 305  
Qy 683 CCAAAGAGAGACAGGCTAATGATCTAGCAGAGACTCAAGGAGAGATTTTCAATTAC 742  
Db 306 CAGCAGCA-----GCCGCAAGAGCTGGCCCGGAGACACAAAGGCCGCTTTCTTGAC 359  
Qy 743 AGAGATGATTTCTTTGTCCTTATTTTGAAGCTATCTCAGAGAGCGAGGAGTCTG 802  
Db 360 CGGAGCGGCTTCTCTTCTCCGTAATCTTGATTAACCTGCGGAGCTTGACGCTGCTG 419  
Qy 803 CCTGATCACTTTCAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 862  
Db 420 CCGACTATCTTCCGAGCGGAGCGCGCTGACGCGCGAGGCGGAGTATTCAGCTGCA 479

Qy 863 GACTGTGTCAACTCTCTGACCCCGAGTGAATCAAGCAAAAGCCAGATGATTTCTGCCAC 922  
Db 480 GAGTCTCTNN 539  
Qy 923 AGTGACTTTGAAGATG 938  
Db 540 NNNNNNTGACAAAG 555

RESULT 7  
US-09-918-995-2311  
Sequence 2311, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: MySeq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
CURRENT APPLICATION NUMBER: US/09/918,995  
PRIORITY FILING DATE: 2001-07-30  
PRIORITY APPLICATION NUMBER: US/09/235,076  
PRIORITY FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2311  
LENGTH: 249  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-918-995-2311

Query Match 2.5%; Score 87; DB 11; Length 249;  
Best Local Similarity 65.2%; Pred. No. 1.4e-10;  
Matches 161; Conservative 0; Mismatches 80; Indels 6; Gaps 2;

Qy 1236 CAGAGTGTGATTTCCACATGTGTGCTGTATCAATCCGTGACAGATCTTTATCAAC 1295  
Db 1 CGAGGCGGCTTCCACATGTGTGCTGTATCAATCCGTGACAGATCTTTATCAAC 60  
Qy 1296 AATATACAGATGACAAAGATCTGTGAAGCTACAGTGAATATGTCTTACCGTGAAGCT 1354  
Db 61 AGTACCGGAGCAAGAGATCTGTGAGCAGCTACACGAGTACATTTTCTTCCGACCACTTC 120  
Qy 1355 --TCCAGATGTGACCCCTCACACTGCGATTGCTGTGCAAGATGGCAAG--GTGACA 1409  
Db 121 AGAAATATGATATCACTTAACAGAACTGATAGATGATACATGACCAAGTCACTGATA 180  
Qy 1410 AAGAAAGGAGAGCGGACAGTGTGCAATGACCTTCCACATCTAGCTGCGAGCGCAAGT 1469  
Db 181 AAGAAAGTGAATGTGGAATCTTCCGATGATGCTTCTTCACTTCCAGTTGGAGACGCAAT 240  
Qy 1470 CTGAGGC 1476  
Db 241 CAGAGGC 247

RESULT 8  
US-10-056-884-8/c  
Sequence 8, Application US/10056884  
Publication No. US20030032786A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL BETA-SUB  
FILE REFERENCE: D0076 NP  
CURRENT APPLICATION NUMBER: US/10/056,884  
PRIORITY FILING DATE: 2002-01-24  
PRIORITY APPLICATION NUMBER: US 60/263,872  
PRIORITY FILING DATE: 2001-01-24  
PRIORITY APPLICATION NUMBER: US 60/269,794  
PRIORITY FILING DATE: 2001-02-14  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: Patent in version 3.0  
SEQ ID NO 8

LENGTH: 80  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic Oligonucleotide Modified To Contain Biotin at the 5 P  
US-10-056-884-8

Query Match 2.3%; Score 80; DB 14; Length 80;  
Best Local Similarity 100.0%; Pred. No. 3.1e-09;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 783 GGGACAGCAGCAGTGGTCTGCTGATCACTTCCAGAAAAAGAGACTGAAGGGAAG 842  
DB 80 GGGACAGCAGCAGTGGTCTGCTGATCACTTCCAGAAAAAGAGACTGAAGGGAAG 21

843 CTGAATCTCTCCAGCTCCCA 862  
CTGAATCTCTCCAGCTCCCA 1

RESULT 9  
US-09-834-975-451/c  
Sequence 451, Application US/09834975  
Patent No. US20020110815A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Brown, Jeffrey  
APPLICANT: Bolt, Andrew  
APPLICANT: Van Hufel, Christophe  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS  
TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
TITLE OF INVENTION: OF HUMAN CANCERS  
FILE REFERENCE: MRI-016B  
CURRENT APPLICATION NUMBER: US/09/834,975  
PRIOR FILING DATE: 2001-04-13  
PRIORITY FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 1046  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 451  
LENGTH: 425  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(425)  
OTHER INFORMATION: n = A,T,C or G  
US-09-834-975-451

Query Match 2.3%; Score 79; DB 10; Length 425;  
Best Local Similarity 50.8%; Pred. No. 1.5e-08;  
Matches 184; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 3107 AATGAACTACTCTTTTCAATACCTGATCTGTATAGACAGCCAAAGAACT 3166  
DB 381 AAAAAAAAAAATATTTTTTTTTTTTTTTTTCACAAAABAAAAAABAAAAAAC 332  
QY 3167 GTTAATAGCAATCCGTCATGTAATCTGTATTTTACAAAGTAGCAATGCTTTTAT 3226  
DB 321 CTTTTTTTTTTTAAAAAGTTTTTTTTTTTAAAAAACCCTTTTGTGGGGG 262  
QY 3227 AGACTGTGCTACAAGAGGTTAAAGACAGTTTATTTTCAGATCTCTGATGCAATTT 3286  
DB 261 GGGGGGATTTTTTTTTTTTTTTTGGAAAAACCCCTTTTTTTTTTTTAAAAAAGC 202  
QY 3287 CAGTGTAAACAAAAAATATTTGTCAATTAATGTTGTGTGCAAGCACTCTTAATTTG 3346  
DB 201 GGGGGGGGAAAAAAGAAAAAACCCTTTTTTTTTTTTTTTTGGGGGAAATTTTTTTTTTT 142  
QY 3347 TTTTATGCGTGTGTGCAATGTTGTATGTATCAAGGTAAATTAAGCAATGATG 3406  
DB 141 TTTTTTTTTTAAAAATTTTTTTTTTTTTTTTNNAAAAAABAAAAAABAAAAA 82

QY 3407 ATTAATAA 3466  
DB 81 AA 22  
QY 3467 AA 3468  
DB 21 AA 20

RESULT 10  
US-09-925-299-112  
Sequence 112, Application US/09925299  
Patent No. US20020055627A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
PRIOR FILING DATE: 2001-08-10  
PRIORITY FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-02-12  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 112  
LENGTH: 1492  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (8)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1487)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1491)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-299-112

Query Match 2.3%; Score 79; DB 9; Length 1492;  
Best Local Similarity 70.2%; Pred. No. 3.3e-08;  
Matches 106; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 3318 ATAGTGTGTGCGCAAGCACTCTTAATTTGTTTATGCGTGTGCGATGTATGT 3377  
DB 1276 AGAATATATATGAGGCAAGATGATGATGATGATGATGATGATGATGATG 1335  
QY 3378 GTATCAGGTATTAAGCAATGATGATTAATAAAAAAAAAAAAAAAAAAAAAA 3437  
DB 1336 AATATATTTTATTAAGAAATTTGATTTCAAAAAAAAAAAAAAAAAAAAAA 1395  
QY 3438 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468  
DB 1396 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1426

RESULT 11  
US-09-925-299-112  
Sequence 112, Application US/09925299  
Patent No. US2003040617A5  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
PRIOR FILING DATE: 2001-08-10  
PRIORITY FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: 60/124,270

NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 112  
; LENGTH: 1492  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (8)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1487)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1491)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-299-112

Query Match 2.3%; Score 79; DB 11; Length 1492;  
Best Local Similarity 70.2%; Pred. No. 3.3e-08;  
Matches 106; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

3318 ATAGTTGTGTCAGCACTCTTAATTGTTTATGCGTGTGTCATGTGTATGT 3377  
1276 AGAATATATTGAGCGCAAGTTCACTTGATGACATGTGTATGTACTGATGCTGT 1335  
3378 GTATCAGGTAATTAAGCAATTGATGATTAATAAAAAAAAAAAAAAAAAA 3437  
1336 AATTATTTTATTAAGAAATTTGATATATCAAAAAAAAAAAAAAAAAA 1395  
3438 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468  
DB 1396 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1426  
QY

RESULT 12  
US-10-106-698-1525  
; Sequence 1525, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: PA005P1  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: Patentin Ver. 3.0  
; SEQ ID NO 1525  
; LENGTH: 1493  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (8)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1488)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc\_feature  
; LOCATION: (1492)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-10-106-698-1525

Query Match 2.3%; Score 79; DB 14; Length 1493;  
Best Local Similarity 70.2%; Pred. No. 3.3e-08;  
Matches 106; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 3318 ATAGTTGTGTCAGCACTCTTAATTGTTTATGCGTGTGTCATGTGTATGT 3377  
DB 1277 AGAATATATTGAGCGCAAGTTCACTTGATGACATGTGTATGTACTGATGCTGT 1336  
QY 3378 GTATCAGGTAATTAAGCAATTGATGATTAATAAAAAAAAAAAAAAAAAA 3437  
DB 1337 AATTATTTTATTAAGAAATTTGATATATCAAAAAAAAAAAAAAAAAA 1396  
3438 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468  
DB 1397 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1427

RESULT 13  
US-10-198-846-8619/c  
; Sequence 8619, Application US/10198846  
; Publication No. US2003009974A1  
; GENERAL INFORMATION:  
; APPLICANT: LiJile, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinhmann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198,846  
; PRIOR FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8619  
; LENGTH: 361  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 15, 21, 25, 26, 29, 30, 33, 50, 59, 242, 243, 245, 247, 249,  
; LOCATION: 266, 268, 279, 281, 282, 283, 285, 286, 287, 289, 290, 296,  
; LOCATION: 297, 311, 312, 313, 324, 325, 326, 329, 331, 343, 344, 351,  
; LOCATION: 352, 353, 361  
; OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-8619

Query Match 2.2%; Score 76.6; DB 14; Length 361;  
Best Local Similarity 54.8%; Pred. No. 5.1e-08;  
Matches 115; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 3259 TTTTATTTAGCATTCCTCATGCAATTCAGTGTAAACCAAAATATTTGCAATTAA 3318  
DB 360 TTTTNNNNAAAAAAAAANCCCCCCCCCGGNGNNNAAAAAAAAAANNNTTTTTTTT 301  
QY 3319 TAGTTGTGTCAGCACTCTTAATTGTTTATGCGTGTGTCATGTGTATGTG 3378  
DB 300 TTTNNGGGGNNNNNNNNNTTTTTTTTTTTNNAAAAAATTTTTNTNTNNA 241  
QY 3379 TATCAGGTAATTAAGCAATTGATGATTAATAAAAAAAAAAAAAAAAAA 3438  
DB 240 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 181  
QY 3439 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468  
DB 180 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 151

RESULT 14  
US-09-770-444-872/c  
; Sequence 872, Application US/09770444  
; Patent No. US20020023280A1  
; GENERAL INFORMATION:  
; APPLICANT: Goriach, Jörn  
; APPLICANT: An, Yong-Qiang

```

/ APPLICANT: Hamilton, Carol M.
/ APPLICANT: Price, Jennifer L.
/ APPLICANT: Raines, Tracy M.
/ APPLICANT: Yu, Yang
/ APPLICANT: Rameaka, Joshua G.
/ APPLICANT: Page, Amy
/ APPLICANT: Matthew, Abraham V.
/ APPLICANT: Ledford, Brooke L.
/ APPLICANT: Moessner, Jeffrey P.
/ APPLICANT: Haas, William David
/ APPLICANT: Garcia, Carlos A.
/ APPLICANT: Kricker, Maja
/ APPLICANT: Slader, Ted
/ APPLICANT: Davis, Keith R.
/ APPLICANT: Allen, Keith
/ APPLICANT: Holman, Neil
/ APPLICANT: Hurban, Patrick
/ TITLE OF INVENTION: Expressed Sequences of Arabidopsis
/ FILE REFERENCE: 2027 (PAPA-016PRV)
/ CURRENT APPLICATION NUMBER: US/09/770,444
/ CURRENT FILING DATE: 2001-01-26
/ PRIOR APPLICATION NUMBER: 60/178,502
/ PRIOR FILING DATE: 2000-01-27
/ NUMBER OF SEQ ID NOS: 999
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 872
/ LENGTH: 442
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-09-770-444-872

```

```

Query Match      2.2%; Score 75; DB 9; Length 442;
Best Local Similarity 64.9%; Pred. No. 1.4e-07;
Matches 11; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 3298 AAAAAAAAAATTTGTCATTAATAGTGTGTCACAGCACTCTAATTTGTTATGCGT 3357
    |||||
Db 172 AAAAAAGGTTGGTATTTATGTTGTCATACATTTCTATTGTAATCTCCG 113

Qy 3358 GTGTGTCATGTGTGTATGTATCAAGGTAATTAAGCAATTGATGATTAATAAAAA 3417
    |||||
Db 112 TTAATGTAATGAAGAGTGTGATGATTAACCAAGTTCAGTTCTTAATAAAAAAA 53

Qy 3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
    |||||
    52 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2

```

```

RESULT 15
/ Sequence 66, Application US/09739254
/ Patent No. US20010021700A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: 49 Human Secreted Proteins
/ FILE REFERENCE: P2032P1
/ CURRENT APPLICATION NUMBER: US/09/739,254
/ CURRENT FILING DATE: 2000-12-19
/ EARLIER APPLICATION NUMBER: 09/511,554
/ EARLIER FILING DATE: 2000-02-23
/ EARLIER APPLICATION NUMBER: PCT/US99/19330
/ EARLIER FILING DATE: 1999-08-24
/ EARLIER APPLICATION NUMBER: 60/097,917
/ EARLIER FILING DATE: 1998-08-25
/ EARLIER APPLICATION NUMBER: 60/098,634
/ EARLIER FILING DATE: 1998-08-31
/ NUMBER OF SEQ ID NOS: 170
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 66
/ LENGTH: 664
/ TYPE: DNA
/ ORGANISM: Homo sapiens

```

```

/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (31)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (63)
/ OTHER INFORMATION: n equals a,t,g, or c
US-09-739-254-66

```

```

Query Match      2.1%; Score 74; DB 9; Length 664;
Best Local Similarity 64.0%; Pred. No. 3.1e-07;
Matches 110; Conservative 1; Mismatches 61; Indels 0; Gaps 0;

Qy 3297 CAAAAAATTTGTCATTAATAGTGTGTCACAGCACTCTAATTTGTTATGCG 3356
    |||||
Db 487 CATAGTGAATTAATTAATTAATGAGGTTTGTCTTTTCTTTTCTTTT 546

Qy 3357 TGTGTGTCATGTGTGTATGTATCAAGGTAATTAAGCAATTGATGATTAATAAAAA 3416
    |||||
Db 547 GGTATATTGCTGTATCTTAACTTCAAGAAATTAAGTTATATRGAAAAA 606

Qy 3417 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
    |||||
Db 607 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 658

```

```

Search completed: September 6, 2003, 11:49:24
Job time : 839 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 5, 2003, 19:09:56 / Search time 6350 Seconds  
(without alignments)  
13273.689 Million cell updates/sec

Title: US-10-056-884-1

Perfect score: 3468  
Sequence: 1 caagcaccgycgaagctc.....aaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_estro:\*  
9: em\_estro:\*  
10: gb\_estc1:\*  
11: gb\_estc2:\*  
12: gb\_estc3:\*  
13: gb\_estc4:\*  
14: gb\_estc5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_ges\_hum:\*  
18: em\_ges\_inv:\*  
19: em\_ges\_pln:\*  
20: em\_ges\_vrt:\*  
21: em\_ges\_fun:\*  
22: em\_ges\_mam:\*  
23: em\_ges\_mus:\*  
24: em\_ges\_dro:\*  
25: em\_ges\_fod:\*  
26: em\_ges\_bhg:\*  
27: em\_ges\_vrt1:\*  
28: gb\_gesb1:\*  
29: gb\_gesb2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	479	13.8	489	28	AQ536411
2	442.4	12.8	810	11	AK015313
3	400.2	11.5	592	28	AQ525390
4	352.2	10.2	784	14	CA463745

Result No.	Score	Query Match	Length	DB ID	Description
C 5	352	10.1	778	13	BU961910
6	350.6	10.1	952	14	BY714867
7	338.6	9.8	489	10	BF391086
8	329	9.5	2332	11	AK043351
9	327.4	9.4	2343	11	AK047519
10	324.2	9.3	2584	11	AK045439
11	322.8	9.3	2555	11	AK042569
12	321.0	8.9	973	13	BO713664
13	305.4	8.8	319	9	AA320222
14	286.2	8.3	424	14	BY708433
15	286	8.2	422	11	AK005368
16	284.2	8.2	419	10	BF413617
17	282.6	8.1	1694	11	AK039167
18	280.6	8.1	401	11	AK005863
19	277	8.0	943	13	BY093872
20	270.2	7.8	400	14	BY706005
21	257.6	7.4	386	10	BF413618
22	255.4	7.4	1710	11	AK046557
23	255.4	7.4	1898	11	AK050097
24	255.4	7.4	3552	11	AK085035
25	255.4	7.4	3572	11	AK046708
26	241.2	7.0	579	12	BU095683
27	241.2	7.0	584	12	BU096172
28	234.8	6.8	491	12	BU095666
29	228.4	6.6	870	14	CA470866
30	220.2	6.3	675	10	BB653208
31	220	6.3	1746	11	AK082563
32	219.2	6.3	295	10	BF522617
33	217.6	6.3	680	10	BA183090
34	217.4	6.3	324	9	AT596442
35	207.2	6.0	433	9	AA497980
36	190.6	5.8	638	12	BM947722
37	188.4	5.5	927	14	CA864920
38	188.4	5.4	709	13	BU294624
39	188.4	5.4	304	9	AV046252
40	186.4	5.4	594	9	AL588086
41	179.2	5.2	310	9	AV206002
42	170.8	4.9	324	9	AV206367
43	169	4.9	295	9	AV040180
44	167.6	4.8	642	12	BU031784
45	167	4.8	1003	29	CNS05PKF

#### ALIGNMENTS

RESULT 1  
A0536411/C  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

489 bp DNA linear GSS 18-MAY-1999  
RPCT-11-318B21.TJ RPCT-11 Home sapiens genomic clone RPCT-11-318B21  
genomic survey sequence.  
A0536411  
A0536411.1 GI:4848101  
GSS  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 489)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter  
J.C.  
Use of BAC End Sequences from Library RPCT-11 for Sequence-Ready  
Map Building  
Unpublished  
Other GSSs: RPCT-11-318B21.TV  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPCT-11. For BAC

library availability. please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACBAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genetics (<http://info.regen.com>). BAC end search page:  
[http://www.flyc.org/cdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.flyc.org/cdb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: Sp6  
Classes: BAC ends.

# FEATURES

Source  
Location/Qualifiers  
1..489  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7621772"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-318B21"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_1b="RPCI-11"  
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
RPC11 Human Male BAC Library"

BASE COUNT 131 a 114 c 105 g 137 t 2 others

Query Match 13.8%; Score 479; DB 28; Length 489;  
Best Local Similarity 98.6%; Pred. No. 3.3e-42;  
Matches 482; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 383 GGATAGAGAGAGGATCTTTTATTAATAGTATGATCCCTTCTTCTTCAAGTTGAT 442  
DB 489 GGATAGAGAGAGGATCTTTTATTAATAGTATGATCCCTTCTTCTTCAAGTTGAT 430  
OY 443 CCAAGGATTAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 502  
DB 429 CCAAGGATTAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 370  
OY 503 AAGAAAGGACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 562  
DB 369 AAGACAGGACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 310  
OY 563 TCCGAGATTCCTCACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 622  
DB 309 TCCGAGATTCCTCACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 250  
OY 623 TTACTCCGCAATTCATGATTAAGATGATGATGATGATGATGATGATGATGATGAT 682  
DB 249 TTACTCCGCAATTCATGATTAAGATGATGATGATGATGATGATGATGATGATGAT 190  
DB 683 CCAAGAGAGACAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 742  
DB 189 CCAAGAGAGACAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 130  
OY 743 AAGATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 802  
DB 129 AAGATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 70  
OY 803 CCTGATCACTTCCAGAAAAGAGACTGAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 862  
DB 69 CCTGATCACTTCCAGAAAAGAGACTGAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 10  
OY 863 GACTTGCTC 871  
DB 9 GACTTGCTC 1

RESULT 2 810 bp mRNA linear HTC 05-DEC-2002  
LOCUS AK015313  
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched  
library, clone:4930434H12, product:inferred, RIKEN cDNA 4930434H12  
gene / putative (Mus musculus), full insert sequence.  
ACCESSION AK015313  
VERSION AK015313.1 GI:12853602  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL  
MEDLINE  
PUBMED  
10349636

## REFERENCE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Komuro, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

## TITLE

JOURNAL  
MEDLINE  
PUBMED  
20499374

## REFERENCE

3 Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,  
Komuro, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multichannel sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

## TITLE

JOURNAL  
MEDLINE  
PUBMED  
20530913

## REFERENCE

4 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Aizawa, T., Hara, A., Fukunishi, Y., Komuro, H., Adachi, J., Fukuda, I.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamazaki, S.,  
Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kusakawa, T., Saito, R.,  
Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, M., Gaasterland, T., Giesi, C., King, B., Kouchiya, H.,  
Kuen, P., Lewis, S., Matsuo, Y., Nishida, C., Suzuki, R., Tomita, M.,  
Quackenbush, J., Schrim, L., Staudt, F., Furum, M., Aono, H.,  
Magneir, U., Washio, T., Sakai, K., Okido, I., Bojunga, N.,  
Baldarelli, R., Baran, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M., F., Brownstein, M., J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gattinich, S., Hill, D.,  
Hofmann, M., Hume, D., A., Kamiya, M., Lee, N., Lyons, P., Nordone, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Momtaz, P.,  
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, R., Suzuki, H.,  
Toyo-oka, K., Wang, K., H., Weitz, C., Whitaker, C., Wilming, L.,  
Wyshaw-Borls, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)

## TITLE

JOURNAL  
MEDLINE  
PUBMED  
21085660

## REFERENCE

5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-575 (2002)

## TITLE

JOURNAL  
MEDLINE  
PUBMED  
11217851

## REFERENCE

6 (bases 1 to 810)  
Aizawa, K., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,  
Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,  
Furum, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,  
Hitoaka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,  
Kawakawa, T., Kato, H., Kawai, J., Kojima, Y., Komuro, H., Kondo, M.,  
Koyama, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,  
Nomura, K., Numata, R., Ono, M., Okazaki, Y., Ohtsuka, T., Owa, C.,  
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y.,  
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, K.,  
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,  
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.







Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>  
 Plate: LICM1080 row: n column: 06  
 High quality sequence stop: 473.

## FEATURES

## source

1 778  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /cd\_xref="taxon:10090"  
 /clone="IMAGE:6742567"  
 /lab\_host="DH10B (TI-phage-resistant)"  
 /clone\_id="NIH\_MGC\_169"  
 /note="Organ: Testicles; Vector: pDRN-LIB; Site\_1: SfiI (ggccatcggcc); Site\_2: SfiI (ggccatcggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:  
 5'-AACGAGTGTATCAGCAGGAGTGGCCTTACGGCCGG-3' and  
 5'-ATTCTAGAGCCGAGCGCGGCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIH, NIH). Note: this is a NIH\_MGC Library."  
 JE COUNT 226 a 180 c 199 g 160 t 13 others

## ORIGIN

## Query Match

10.1%; Score 352; DB 13; Length 778;  
 Best Local Similarity 85.6%; Pred No. 8.9e-29;

## Matches

404; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

## ORIGIN

## Query

1347 GTGAGCTTCCAGATGTCACCTTCACCTGCGATGTCGCAAGAAAGGCAAGG 1406

## Db

183 GTGAGCTTCCGCGGTGTCCTCTCATTTGATGACCTGCGCAAGAAAGGCAAGG 242

## Query

1407 ACGAAGAGGAGAGAGCGGACCGTCTTGCAATGACCTTCCACATTTAGCTCGACGCC 1466

## Db

243 ACA--AGGAGAGAGAGGCGACCTCTCGCAATGACCTTCCACATTTAGCTCGACGCC 299

## Query

1467 AGTGTAGAGGCGAGCTCTCCCGAGAGACCGTCTGATCTGTGGGCTGTAAACGGCCAGAGCA 1526

## Db

300 AGTCAGAGGCGAGCTCTCCCGAGAGACCGTGTATCTGTGGGCTGTAAACGGCCAGAGCA 359

## Query

1527 ACATCCAGACTGTGACCGTCCCATCAAGAGAGGCGCTGTCCAGCTGATCCACATGTGAG 1586

## Db

360 ACATCCAGACTGTGATGCGCCATCAAGAAAGCTCGGTGACGCTATCCACATGTGAG 419

## Query

1587 AGATGGCGCGGAAAGGCACTTACTCCGATTTCTGCTTCAAGGCTCCAGGAATCCGAACA 1646

## Db

420 AGATGGCGCGGAAAGGCACTTACTCCGATTTCTGCTTCAAGGCTCCAGGAATCCGAACA 479

## Query

1647 TGAAGCAAGCAAGAAAGCTTTAAAGAAAGCTCTCAATTGAGAGAGAGCTGAGAAAT 1706

## Db

480 TGAAGCAAGCAAGAAAGCTCTGCAAGAAAGCTCTCTCATCGAGAGAGAGCTGAGAAAT 539

## Query

1707 GTATCCAGAGATTTCTTAATAAATAATTCACATGCTTCTTCTGAGAGAAACATCTTT 1766

## Db

540 GTATCCAGAGATTTCTTGAAGATTAATAATTCACATGCTTCTTCTGAGAGAAACATCTTT 599

## Query

1767 GGCATCTGAACCTTTTAAAGAGATATCATATATAGGAGAGGCTGGGGCGG 1818

## Db

600 GGCAGTCTGAACCTTTTACGAGATATCATATATAGGAGAGGCTGTGGGG 651

RESULT\_6  
 LOCUS BY14867  
 DEFINITION BY14867 RIKEN full-length enriched, adult male testis Mus musculus  
 CDNA clone 4930434H12 5', mRNA sequence.  
 ACCESSION BY14867  
 VERSION BY14867.1 GI:27127984  
 KEYWORDS EST  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

## REFERENCE

1 (bases 1 to 952)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nishikawa, I., Otsu, N., Saito, R., Suzuki, H., Yamanka, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.C.,  
 Quackenbush, J., Schmitt, L.M., Kanpin, A., Matsuda, H., Batalov, S.,  
 Baisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chochia, C., Corbett,  
 L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Foxrest,  
 A., Frazer, K.S., Gasstrel, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,  
 Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M.,  
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lemnar, B., Lyons,  
 E.A., Maglott, R., Maltsev, L., Marchionni, L., McKenzie, L., Miki,  
 P., Nagashima, T., Numata, K., Okido, T., Pavani, W.J., Petosa, G.,  
 Pesole, G., Petrovsky, N., Pili, A., Reed, J.C., Reed, D.J., Reid, J., Ring,  
 Ramchandran, S., Ravasi, I., Reed, J.C., Reed, D.J., Reid, J., Ring,  
 B.Z., Ringwald, M., Sander, A., Schneider, C., Sepp, C.A., Setou,  
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,  
 R.D., Tomita, M., Verardo, R., Wagner, L., Walstedt, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L.G., Wymshaw-Boris, A., Yanagisawa,  
 M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A.,  
 Carninci, P., Hayatsu, N., Hirozane-Kitahara, T., Konno, H., Nakamura,  
 M., Sakazume, N., Sato, K., Shizaki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
 K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
 E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

## TITLE

Analysis of the mouse transcriptome based on functional annotation

## JOURNAL

Nature 420, 563-573 (2002)

## MEDLINE

12466851

## PUBMED

12466851

## COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsukuba-shi, Ibaraki, 305-3858, Japan  
 Tel.: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-resseq@riken.go.jp/  
 URL: <http://genome.gsc.riken.go.jp/>  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda,  
 S., Hashizume, M., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,  
 Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno,  
 H., Koye, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,  
 Numazaki, R., Ono, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H.,  
 Sasaki, D., Sato, K., Shibata, K., Shizaki, T., Tagami, M., Takeda, Y.,  
 Waki, K., Watanabe, A., Watanabe, M. and Hayashizaki, Y. Direct  
 Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN Integrated sequence analysis (RISA) system--384-Gomat  
 sequencing pipeline with 384 multiplexed sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
 further details.

## FEATURES

## source

Location/Qualifiers  
 1. .952  
 /organism="Mus musculus"  
 /mol\_type="mRNA"

	/strain="C57BL/6J"	
	/db_xref="taxon:10090"	
	/clone="4930434H12"	
	/sex="male"	
	/tissue_type="testis"	
	/dev_stage="adult"	
	/lab_host="SOLR"	
	/clone_idb="RIKEN full-length enriched, adult male testis	
	/note=Site 1: XhoI; Site 2: BamHI. cDNA library was	
	prepared and sequenced in Mouse Genome Encyclopedia	
	Project of Genome Exploration Research Group in Riken	
	Genomic Sciences Center and Genome Science Laboratory in	
	RIKEN. Division of Experimental Animal Research in Riken	
	contributed to prepare mouse tissues. 1st strand cDNA was	
	primed with a primer [5']	
	GAGAGGAGGAGGATCCACAGACTCTTTTTCCTTTTTTTTNN 3'), cDNA was	
	prepared by using trehalose thermo-activated reverse	
	transcriptase and subsequently enriched for full-length by	
	cap-trapper. Second strand cDNA was prepared with the	
	primer adapter of sequence [5']	
	GAGAGGAGGAGCGGCCCATTAATTATTCCTGAGTTAATAATATCCCCCCC	
	3'). cDNA was cloned into the XhoI and BamHI sites. "	
BASE COUNT	223 a 259 c 226 g 229 t 5 others	
ORIGIN		
Query Match	10.1%; Score 350.6; DB:14; Length 952;	
Best Local Similarity	75.5%; Pzed No. 11e-28;	
Matches 506; Conservative 0; Mismatches 142; Indels 22; Gaps 5;		
Dy	1347 GTGAGCTTCCAGATGTGCACCTCCACATGCTGCTGTCGAAGAATGGCAAAAGTG 1406	
Dd	167 GTGAGCTTCCCGATGATCTCTCTTCAATGTATCTGCTGTCGAAGAATGGCAAAAGTG 226	
Dy	1407 ACAAGAGAGGAGAGCGGACGCTTTCATATGACCTCTCCACATCTAGTGGCACAGCC 1466	
Dd	227 ACA--AAGAGAGAGAGCGGACCTCTTCATATGACCTCTTCACATCTTAGTGGCACAGCC 283	
Dy	1467 AGTCTGAGGCGACCTCTCCCGAGAGAGGTCATCTGTGTCCCGTGCACGCGCACAGCA 1528	
Dd	284 AGTCAGAGGCGCACCTCTCCGAGAGAGAGGTCATCTGTGTGCGCTGTAAACGCGCACAGCA 343	
Dy	1527 ACATCCAGACTCTGAGACCGTCCCATCAAGAAAGGCCCTCTCCAGCTGATCCACAGCTGAG 1586	
Dd	344 ACATCCAGACTCTGAGATCGGCCCATCAAGAAAGGTCGGTGAAGTATCCACAGCTGAG 403	
Dy	1587 AGATCGCGCGGAAAACGACCTTACCTCCGAGATCTGACCTCAGCTCCAGGAAATCCAGACA 1646	
Dd	404 AATATAGCGGAAAAGTACCTGCTCCGAGATCTGACCTCAGCTCCAGGAAATCCAGACA 463	
Dy	1647 TGAGCAGCAAAAAAAGCTGTAAAGAAAAGCTCTCAATTGAGAGAGAGCTGAGAAAT 1706	
Dd	464 TMAAGCAGCAAAAAAAGAGCTGCGAAGAAAAGCTCTCAATGACAAAGAGCTGAGAAAT 523	
Dy	1707 GTATCCAGAGATTTCTTAAAAAATAATTTCCAGATCGATTTCCTGAGAGAAAATCTCTT 1766	
Dd	524 GTATCCAGAGATTTCTTGAAGATTAATAATTTCCAGATCGATTTCCTGAGAGAACATCTCTT 583	
Dy	1767 GGCAATCTGAACTTTTAAAGAGTATCATATTAAGGAGGCTGGGGCGCGGAGAAAAA 1826	
Dd	584 GGCAATCTGAACTTTTACGGAGTATCATATTAAGGAGGCGGAGCGCGG----- 635	
Dy	1827 AAAAAAAAAAGCTCATTTTGAATTAACCTCAAAAAAGAAATCATATTTTAAAGAAAAA 1886	
Dd	636 -----CTCGGCTCTCAATTAACCCNCTCAAAGGGGAGCANITCTTAAACG-GAAA 686	
Dy	1887 AATTAAGTAAATGATGACATTTCTTAAACAGCAATATGTCATTAATTAATCTGCTTA 1946	
Dd	687 AGTATCACTTAATGANTATCACTCTGACAGCACAGAG-CATCTATATATCTTCTCTTA 745	
Dy	1947 CTTTACCTTAATGATCACTTAACTGAATTCACAGGATGATTTCTTTCTGATGTGGAA 2006	
Dd	746 CTTTATGCTA-CCCTCTTAAAGTGCATCTCAAGGGGACATTTTTTATATGTGATCA 804	

FEATURES	source	Location/Qualifiers
1. 489		
/organism="Rattus norvegicus"		
/mol_type="RNA"		
/strain="Sprague-Dawley"		
/db_xref="taxon:10116"		
/clone="UI-R-CA1-bcd-a-05-0-UI"		
/lab_host="DH10B (Life Technologies)"		
/clone_1ib="UI-R-CA1"		
/note="Vector: pRT3D-Poc I; Pharnacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CA1 library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at <a href="http://ratist.eng.uiowa.edu">ratist.eng.uiowa.edu</a> . The subtraction has been previously described in (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996)		
TAG LIB=UI-R-CA1		
TAG TISSUE=testis		
TAG SRO=ACCGAC		
BASE COUNT	89 a	123 c 111 g 164 t 2 others
ORIGIN		
Query Match	9.8%	Score 338.6; DB 10; Length 489;
Best Local Similarity	84.0%	Pred. No. 3.1e-27;
Matches 400;	Conservative	0; Mismatches 60; Indels 16; Gaps 14
1413	AGGAGAGACGGCAGCTGTTCATATGACTCTTCACATCTAGCTGACACGCGCATCTG	1472



```

/codon_start=1
/db_xref="gi:26315633"
/translation="MALKONGSGSTLLPISEMWASASSSPGAPLAAPGPCAPSPPE
VVEINAGGVYVYVSTLLSVPSSTLSPSSPSPGAPRBDLPRDRAFPDMD
GFLPRVYDYLRDROLAPHPPEKRLIPRAEPFLDLYLSPKTKONLNDSC
COSDIEDVVSOSSDALLRGLAAGPSSGASGVSGAPRSGELTGYG
STTVFNDQDAKTRVARTVCGRLAALKEVFGTILABERDPRPEYTRRYLKE
THEDKRLKVDKSGESGTSNLSSTSDSHSASTFSPDNPATQAAHPVSPKO
EHEDRKRVDSKESGTSNLSSTSDSHSASTFSPDNPATQAAHPVSPKO
DPRKAPVQWPPDPDRKRSSELFSLSKSEFTLSKRVCEKLSVEEMKCIQDP
KHLIDPEPRKQWSELQRYGL"
BASE COUNT      556 a      647 c      620 g      509 t
ORIGIN
Query Match      9.5%; Score 329; DB 11; Length 2332;
Best Local Similarity 58.5%; Pred. No. 1,3e-26;
Matches 781; Conservative 0; Mismatches 455; Indels 99; Gaps 8;

```

```

Db 1261 GCGTCAACTCTCTGCGACTGCGCGCTTTGTCAACGAGACGACGAAGATCTGG 1320
Qy 1319 TCAGCTCACTGAATATGCTTCTACCGTAGCCT---TCAGATAGTACACCTCACAC 1375
Db 1321 AGAGTTACATGAAATACATCTTCTTCGACCACTCAGAAAATAGTGTCACCAAGCA 1380
Qy 1376 TCGATTCCTCTCGAAGATGCGAAG---GTGACAAAGAGGAGAGCGGACGCT 1432
Db 1381 GAACATGAAGACAGAAACCGCAAAATGCACAGCAAAAGAAATGAGTGGACTTCC 1440
Qy 1433 TCGATGACTCTTCCACATCTTACCTGCGACACCGAGTGTAGGCGAGCTTCCCGACAG 1492
Db 1441 TCGATGAGCTCTCCACATCTGAGAGCCACCTGACAGGCGACGACATCCACAGAC 1500
Qy 1493 A-----CGTCACTGTGTGTCCTCGTACAGCCGACCAATCATCAAGATCTGAGCGT 1546
Db 1501 AACCCAGCCACACTCAGACAGGTGAGTCCACAGCTTACACTTACCTTGGATGAA 1560
Qy 1547 CCCATCMAAGAGGCGCTGTCTCAGCTGATCCAAAGTCAAGATGCGCGGAGAAAGCGAC 1606
Db 1561 CCTCCAGGAAAGACCTGTTCACTGATGACCAACAGCAAGCAAGCAAAAGTGA 1620
Qy 1607 TTACTCCGATTTCTGACTTCACTGAGCTCCAGGAAATCGAATGACAGCAAAAGAAAGCT 1666
Db 1621 CTCTTCACTCATCTCATCAGCAAGTCCGAGAAACAAATCT-----CTCAAAAAGAA 1674
Qy 1667 GTTAAAGAAAGCTCTCAATTTAGAGAGAGCTGAGAAATGATCCAGATTTCTTAA 1726
Db 1675 GTCTGAGAAAGCTTCAAGTGAAGAAAGAAATGAAATGATATTCAGATTTTAA 1734
Qy 1727 AAAAAATTCAGATCGTTTCTCTGAGAGAAACATCTCTGAGCAATCGAATTTAAAG 1786
Db 1735 ATCCATATTCAGATGTTGTTTTCAGAGCGCAACCCGAGTGAATCTTAACTCTCCA 1794
Qy 1787 AAGTATCATATTA 1801
Db 1795 AATATGAGGTGTAA 1809

RESULT 9
AK047519
LOCUS
DEFINITION
AK047519 2343 bp mRNA linear HTC 05-DEC-2002
Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:B9308201 product:hypothetical protein,
full insert sequence.
ACCESSION
AK047519
VERSION
AK047519.1 GI:26092232
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Mizumatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE
PUBMED
11042159
11042159
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishii, K., Kikunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M.,
Itoneda, Y., Ishikawa, F., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

```

TITLE	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
	RIKEN Integrated sequence analysis (RISA) system 384-format		
	Sequencing pipeline with 384 multiplexed sequencer		
	Genome Res. 10 (11), 1757-1771 (2000)		
JOURNAL	20030913		
PUBLISHED	11076861		
REFERENCE	4		
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,		
	Aizawa, T., Hara, A., Fukunishi, Y., Komoto, H., Adachi, J., Fukuda, S.,		
	Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamashita, I.,		
	Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kanakawa, T., Saito, R.,		
AUTHORS	Kadoya, K., Matsuda, H., Ashburner, M., Baldo, V., Casavant, T.,		
	Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochwa, H.,		
	Kuehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pesole, G.,		
	Quackenbush, J., Schriml, L.M., Staudt, J., Suzuki, R., Tomita, M.,		
AUTHORS	Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,		
	Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,		
	Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Buit, C.,		
	Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,		
AUTHORS	Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Nordone, P.,		
	Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P.,		
	Rung, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,		
	Sato, K., Schenbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,		
AUTHORS	Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Williams, L.,		
	Watanabe, B., Wang, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.,		
	and Hayashizaki, Y.		
	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409 (6821), 685-690 (2001)		
PUBLISHED	21085660		
REFERENCE	11217851		
AUTHORS	5		
	The FANTOM Consortium and the RIKEN Genome Exploration Research		
	Group, Phase I & II Team.		
	Analysis of the mouse transcriptome based on functional annotation		
JOURNAL	of 60,770 full-length cDNAs		
PUBLISHED	Nature 420, 563-573 (2002)		
REFERENCE	6 (Passes 1 to 2343)		
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P.,		
	Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,		
	Hayashida, K., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hirose, T.,		
	Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kageura, T.,		
AUTHORS	Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Kondo, M.,		
	Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,		
	Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnishi, N.,		
	Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,		
AUTHORS	Sato, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,		
	Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,		
	Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,		
	Muramatsu, M., and Hayashizaki, Y.		
AUTHORS	Direct Submissions		
	Submitted (16-JUN-2001) Yoshitake Hayashizaki, The Institute of		
	Physical and Chemical Research (RIKEN), Laboratory for Genome		
	Exploration Research Group, RIKEN Genomic Sciences Center (GSC),		
JOURNAL	RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsunumi-ku, Yokohama,		
PUBLISHED	Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,		
REFERENCE	URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,		
AUTHORS	Fax: 81-45-503-9216)		
	CDNA library was prepared and sequenced in Mouse Genome		
	Encyclopedia Project of Genome Exploration Research Group in Riken		
	Genomic Sciences Center and Genome Science Laboratory in RIKEN.		
AUTHORS	Division of Experimental Animal Research in Riken contributed to		
	prepare mouse tissues.		
	Please visit our web site for further details.		
	URL: http://genome.gsc.riken.go.jp/		
JOURNAL	URL: http://fantom.gsc.riken.go.jp/		
PUBLISHED	Location/Qualifiers		
REFERENCE	1 2343		
AUTHORS	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6J"		
	/db_xref="FANTOM_DB:B930082J01"		
AUTHORS	/db_xref="taxon:10090"		
	/clone="B930082J01"		

misc_feature	/tissue_type="cerebellum"		
	/clone_lib="RIKEN full-length enriched mouse cDNA library"		
	/dev_stage="10 days neonate"		
	1.2343		
BASE COUNT	/note="Hypothetical protein (evidence: rcds,Proctree,decoder,Longest-ORF)"		
	554 a 660 c 624 g 505 t		
	ORIGIN		
	Query Match 9.4%; Score 327.4; DB 11; Length 2343;		
Matches	Best Local Similarity 58.4%; Pred. No. 26-26;		
	Matches 780; Conservative 0; Mismatches 456; Indels 99; Gaps 8;		
QY	560	GGGTCGCGACGTTCCCACTCTTCCGAGAGGTGAGAGCTGAATGTGGGGGTCAAGT	619
DB	512	GGGCGCTGGCGCCCGTGGCCCTTCCCGAGAGTGAATGTGGCGCGCAAGT	571
QY	620	TATTTACCGGCATTCACATTTGAATGATTCCTTCCCGTGGAGAAATGTT	679
DB	572	TATGTACCAAGATTGACCTTACACGCTCCGACAGACTGTGGCAATGTT	631
QY	680	TCCCAAGAGAGAGAGAGCTAATATC-----TAGCAAGATCCCAAG	724
DB	632	TACCTCTAGTCCCGGGGGGGGGAGAGCCCGGACCTTCCGAGAGAGCCG	691
QY	725	GGAAGTTTTCATTCAGAGATGATTTCTTCTTATTTGACTATCTCAG	784
DB	692	GCGGCTTCTTATGACCGGAGGCTTCTTATGATGATGATTAATCTGCG	751
QY	785	GACAGAGAGTGTCTGCTGATTCATCTTCCAGAAAGAGAGAGAGAGT	844
DB	752	GACAGAGAGTGTCTGCTGATTCATCTTCCAGAAAGAGAGAGAGT	811
QY	845	GAATTTCTGAGTCCAGAGTGTGCAAGTCTGAGAGAGAGAGAGAGT	904
DB	812	GAATTTCTGAGTCCAGAGTGTGCAAGTCTGAGAGAGAGAGAGT	871
QY	905	CC-----AGATGAATTTCTGCAAGTCTGAGAGAGAGAGT	936
DB	872	TCCCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	931
QY	937	-----TCCCTCAAGAGAGAGAGAGAGAGAGAGAGAGT	977
DB	932	GAGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	991
QY	978	CCCTG-----CTCCCTGCGAGCGCAAGTGGGTTTCAATCTGTG	1018
DB	992	GTCAGTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1051
QY	1019	GTTTACAGAGATCTGACCTTGGCAGAGAGAGAGAGAGAGAGT	1078
DB	1052	GCTTACCTGCTCTTACACAGAGAGAGAGAGAGAGAGAGAGT	1111
QY	1079	GTTTCCCGGATTTGTTTGGAGAGATTTCTTGGCAAGAGTCTTTGGAGAACT	1138
DB	1112	GTTGCGCGCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1171
QY	1139	TTGAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	1198
DB	1172	CTTAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	1231
QY	1199	AAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	1258
DB	1232	AAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	1291
QY	1259	GCTTGAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	1318
DB	1292	GCTTGAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	1351
QY	1319	TCAAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	1375
DB	1352	AGCAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT	1411

QY 1376 TGGATTGCTGCTGCAGAGATGCGAAG---GTGCAAGAAAGGAGGAGCGGCACTCT 1412  
 DB 1412 GAACATGAGACAGGAAACGGCAAAAGTCAACAGACGACCAAGAGATGAGTGGACTTCC 1471  
 QY 1433 TCGAATGACCTCTCCACATCTAGCTGAGCAGCAGCTGAGGCGAGCTCTCCAGAG 1492  
 DB 1472 TCGAATGACCTCTCCACATCTAGCTGAGCAGCAGCTGAGGCGAGCTCTCCAGAG 1531  
 QY 1493 A-----CGTCATCTGTGCTCCCTGACACCGCAGACCAATCCAGATCTGACCGT 1546  
 DB 1532 AACCCAGCCACACCTGACAGGCTGACAGCTGACAGCTGACAGCTGACAGCTGAC 1591  
 QY 1547 CCCATCAAGAGGCGCTGCTGACCTGATCAACAGTCAAGATGCGCGGAGAACGAC 1606  
 DB 1592 CCTCCAGAGAAAGCAGCTGCTGATGATCCACACCAAGAGGCGAGCAAGTGA 1651  
 QY 1607 TTACTCCGATTTCTGACTTCAAGCTCCAGGGAATGAAATGACAGCAAAAAAGCT 1666  
 DB 1652 CTCCTTCAGTCACTCATGACGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 1705  
 QY 1667 GTTAAAGAAAAGCTCTCAATTCAGAGAGAGCTGAGAGAAATGATTCAGAGATTTCTA 1726  
 DB 1706 GTCTGTAGAGAGCTAGAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1765  
 QY 1727 AAAAAATTCAGAGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786  
 DB 1766 ATCCATATTCAGAGATTTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1825  
 QY 1787 AAGTATCATCTATTA 1801  
 DB 1826 AATATTCGCTGTAA 1840

RESULT 10  
 AK045439  
 LOCUS  
 DEFINITION  
 AK045439 2584 bp mRNA linear HTC 05-DEC-2002  
 Mus musculus adult male corpora quadrigemina cDNA, RIKEN  
 full-length enriched library, clone:B3J0119K12 product:hypothetical  
 protein, full insert sequence.  
 AK045439  
 VERSION  
 AK045439.1 GI:26337364  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 99279253  
 10349636  
 REFERENCE  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 20493574  
 11042159  
 REFERENCE  
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
 Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 20530913  
 11076861

REFERENCE  
 AUTHORS  
 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,  
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
 Fleischmann, W., Gasterland, T., Gissi, C., King, B., Kochiya, H.,  
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pasale, G., Koshita, H.,  
 Quackenbush, J., Schriml, L. M., Stenbali, P., Suzuki, R., Tomita, M.,  
 Wagner, E., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
 Balaguer, E., de Bonaldo, M. F., Brownstein, M. J., Bul, C.,  
 Carninci, P., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
 Holman, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
 Marchionni, L., Mashima, U., Mazzarelli, V., Mombaur, P., Nordone, P.,  
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, R.,  
 Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,  
 Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohlschki, S.  
 and Hayashizaki, Y.  
 TITLE  
 Functional annotation of a full-length mouse cDNA collection  
 JOURNAL  
 MEDLINE  
 PUBMED  
 Nature 409 (6821), 685-690 (2001)  
 21085660  
 11217851  
 REFERENCE  
 5  
 TITLE  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 2584)  
 REFERENCE  
 6  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hayashizaki, Y.,  
 Hayashida, K., Hayatsu, N., Hiyamori, K., Hirooka, T., Hiyama, T.,  
 Hori, F., Imocanu, K., Ishii, Y., Itoh, M., Kasegawa, T., Kasegawa, T.,  
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Konda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Nakazaki, R., Ono, M., Onose, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 COMMENT  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/  
 Location/Qualifiers  
 1. 2584  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM:DB:B3J0119K12"  
 /db\_xref="taxon:10090"  
 /clone="B3J0119K12"  
 /sex="male"  
 /tissue="corpora quadrigemina"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 262...1692  
 /note="unlabeled protein product; hypothetical protein

CDS



TITLE	Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wachihi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
	RIKEN Integrated sequence analysis (RISA) system-384-format	
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE	20530913	
PIRME	11076661	
REFERENCE	4	
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Atakawa, I., Hara, A., Fukunishi, I., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishikawa, K., Kiyosawa, H., Kondo, S., Tanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Cavaletto, F., Fleschmann, W., Gaasterland, T., Glast, C., King, B., Kuchitsu, H., Kuehl, P., Lewis, S., Matsuo, Y., Mikado, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bolunga, N., Carinici, P., de Bona, M. P., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringuwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitlaker, C., Wilmberg, L., Wyszynski, B., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.	
TITLE	Functional annotation of a full-length mouse cDNA collection	
	Nature 409 (6821), 685-690 (2001)	
JOURNAL	21085660	
MEDLINE	1217851	
PIRME	5	
REFERENCE	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II team.	
AUTHORS	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
JOURNAL	Nature 420, 563-573 (2002)	
MEDLINE	6 (bases 1 to 2555)	
PIRME	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carinici, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	
TITLE	Direct Submission	
	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-research.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	
JOURNAL	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
COMMENT	Please visit our web site for further details.	
FEATURES	URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers	
source	1. .2555 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM_DB:A730006K23"	

COS

/db\_xref="taxon:10090"  
/clone="A730006K23"  
/tissue\_type="cerebellum"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="7 days neonate"  
262..1695  
/note="unnamed protein product; hypothetical protein  
(evidence: rcds, ProCreast, decoder, Longest-ORF)  
putative"  
/codon\_start=1  
/protein\_id="BA31296.1"  
/db\_xref="GI:26335191"  
/translation="MLKRDGSGSTLIPISWVSSSPGAPLIAAGPCAPSPPE  
IVRLNGCQVYVYKHTLISVDPSTLAMPSPSPGAPRPGLPDPDSAPFIR  
DGFPRVLDIADNOLPPEPEKPLAPAPFQDLDYKLSPTVYQNSLINS  
CCQDLSDNVSQSSDALLRLRAAAGPSGAGAGVSGVGGSAADPRSGFTLGYR  
GSYTVLDNDKRRFVRIKIVCRITLAKVPGDTLNSDPRDPEKTSFYK  
FTYLEQAFRLISAGHVAACSGTAFVQTRDRLNSSTETITFRPQKIVSK  
QEHDRDRDVTVDSESTSCNLSSTSCDSHSEASTPDNPATQQAHAHQNTIL  
LDRSRAPVQWMPDPKRNSELPQSLISRETNLSKKKVCERLSEEMKKTID  
FKIHIDCFPERKQWSEELQKYL"

BASE COUNT

666 a 638 c 643 g 608 t

ORIGIN

Query Match

Best Local Similarity 58.2%; Score 322.8; DB 11; Length 2555;

Matches 779; Conservative 0; Mismatches 457; Indels 102; Gaps 8;

560 GGGTCGGAGTTCCTCCATCTCTCTAGGTGTAGAGTGAATGCGGGGCTCAAGTT 619  
619 GGGTCGGAGTTCCTCCATCTCTCTAGGTGTAGAGTGAATGCGGGGCTCAAGTT 619  
364 GGGTCGGAGTTCCTCCATCTCTCTAGGTGTAGAGTGAATGCGGGGCTCAAGTT 423  
423 GGGTCGGAGTTCCTCCATCTCTCTAGGTGTAGAGTGAATGCGGGGCTCAAGTT 423  
620 TATTTTATCTCCGATTCATGATGATGATGATGATGATGATGATGATGATGAT 679  
679 TATTTTATCTCCGATTCATGATGATGATGATGATGATGATGATGATGATGAT 679  
424 TATTTTATCTCCGATTCATGATGATGATGATGATGATGATGATGATGATGAT 483  
483 TATTTTATCTCCGATTCATGATGATGATGATGATGATGATGATGATGATGAT 483  
680 TCCCAAGAGACAGGCTATGATC-----TAGCAAGAGATCC 721  
721 TCCCAAGAGACAGGCTATGATC-----TAGCAAGAGATCC 721  
484 TCACTCTGATGCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 543  
543 TCACTCTGATGCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 543  
722 AAGGAGGATTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGAT 781  
781 AAGGAGGATTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGAT 781  
544 CGGAGCGGCTTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGAT 603  
603 CGGAGCGGCTTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGAT 603  
782 AAGGAGGATTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGAT 841  
841 AAGGAGGATTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGAT 841  
604 CGGAGCGGCTTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGAT 663  
663 CGGAGCGGCTTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGAT 663  
842 GCTGAATTTCTCCAGCTCCAGATGATGATGATGATGATGATGATGATGATGAT 901  
901 GCTGAATTTCTCCAGCTCCAGATGATGATGATGATGATGATGATGATGATGAT 901  
664 GAGAGGTTTCTCCAGCTCCAGATGATGATGATGATGATGATGATGATGATGAT 723  
723 GAGAGGTTTCTCCAGCTCCAGATGATGATGATGATGATGATGATGATGATGAT 723  
902 AGCC-----AGATGAATTTCTCCAGATGATGATGATGATGATGATGATGAT 936  
936 AGCC-----AGATGAATTTCTCCAGATGATGATGATGATGATGATGATGAT 936  
724 AACTGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 783  
783 AACTGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 783  
937 -----TGCTCCAGAGAGAGACAGACAGAGATTTGCCCCCTT 974  
974 -----TGCTCCAGAGAGAGACAGACAGAGATTTGCCCCCTT 974  
784 AGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 843  
843 AGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 843  
975 CTCTCTCTG-----CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1015  
1015 CTCTCTCTG-----CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1015  
844 GGTGTAGGT 903  
903 GGTGTAGGT 903  
1016 GTGGGTTTACAGAGATCTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075  
1075 GTGGGTTTACAGAGATCTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075  
904 CTGGGTTTACAGAGATCTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 963  
963 CTGGGTTTACAGAGATCTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 963  
1076 AGAGTTTCTCCAGATTTTGTGAGAGAGATTTTGTGAGAGAGAGAGAGAGAG 1135  
1135 AGAGTTTCTCCAGATTTTGTGAGAGAGATTTTGTGAGAGAGAGAGAGAGAG 1135  
964 CGTGTGGGCGGATCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1023  
1023 CGTGTGGGCGGATCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1023  
1136 ACTTGAATGAAG 1195  
1195 ACTTGAATGAAG 1195





```

/organism:"Homo sapiens"
/mol_type:"mRNA"
/db_xref:"PRTC (inhet):133650"
/db_xref:"taxon:9606"
/dev_stages:"embryo, 8 wks"
/clone_lib:"Embryo, 8 week I"
name:"Organ: Embryo, 8 weeks"
Vector: pbluescript SK-;

```

Yaglik, R., Tomari, T., Hasegawa, T., Nogami, A., Schumacher, C., Gojobori, T., Balderelli, R., Hill, D. P., Bult, C., Hume, D. A., Koike, K., Kuroki, Y., Schriml, L. M., Kampan, A., Matsuda, H., Batalov, S., Beisel, K. M., Blake, J. A., Brad, D., Breich, V., Chonka, C., Corbett, L. B., Cousins, S., Daille, E., Dargatzis, A., Fletcher, C., Foster, A., Fraser, R. S., Gaasterland, T., Gaidarov, I., Gao, J., Gao, J. A., Gough, D., Gilmann, S., Gostencnik, S., Hayakawa, N., Jackson, J. A., Karas, E. D., Konata, A., Kowal, R. J., Kras, S., Kozlov, L. R., Kozlov, R. M., Kung, B., Kung'u, A., Lalancette, J., Lee, Y., Leifert, B., Lyons, P. A., Mchugh, T. R., Numata, K., Okido, T., Pavani, W. J., Petrea, G., H. A. Noshay, T. R., Numata, K., Okido, T., Pavani, W. J., Petrea, G., Peleg, E., Petrovsky, N., Pillai, R., Reed, J. C., Reed, J. U., Qi, D., Ramchandani, S., Ravasi, T., Reed, J. C., Reed, J. U., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Selton, M. Z., Shinada, K., Sultana, R., Takemura, C., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wablstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, I. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, L., Yuan, Z., Zavalon, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirotsane-Tsukikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Waki, J., Aizawa, K., Aizawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Iehi, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

TITLE









[illegible]





Db 2260 CATTGAGAGAGACGTGAGAAATGATTCAGAGATTTCTTAAATTCAGATTCAGATTC 2319  
 Qy 1743 GGTTCCTGAGAGAAACATCTTGGCAATCTGAACTTTTAAAGAGATCATCTATAAG 1802  
 Db 2320 GGTTCCTGAGAGAAACATCTTGGCAATCTGAACTTTTAAAGAGATCATCTATAAG 2379  
 Qy 1803 GGAAGGCTGGGGCGGGGAAAAA 1834  
 Db 2380 GGAAGGCTGGGGCGGGGAAAAA 2411  
 RESULT 3  
 AAD46125  
 ID AAD46125 standard; DNA; 769 BP.  
 AC AAD46125;  
 XX  
 XX 27-DEC-2002 (first entry)  
 DE Human BAC AC008652 exon used to isolate K+betaM2 cDNA.  
 YX  
 Human: potassium channel beta-subunit; K+betaM2 protein; neural disorder;  
 reproductive disorder; metabolic disorder; premature puberty; nephritis;  
 endocrine disorder; memory disorder; neuroendocrine condition; asthma;  
 spermatogenesis; renal disease; learning deficiency; Alzheimer's disease;  
 neurodegenerative disease; proliferative disorder; autoimmune disease;  
 carcinoid tumour; blood coagulation disease; blood platelet disease;  
 rheumatoid arthritis; allergy; hyperproliferative disease; gene therapy;  
 graft-versus-host disease; organ rejection; antileishmaniasis; thrombolytic;  
 anti-inflammatory; neuroprotective; anti-Parkinsonian; immunosuppressive;  
 nephrotropic; cytoskeletal; noctropic; hypotensive; vulnerary; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 PN WC0200266601-A2.  
 XX  
 XX 29-AUG-2002.  
 PF 24-JAN-2002; 2002MO-US02332.  
 XX  
 XX 24-JAN-2001; 2001US-263872P.  
 PR 14-FEB-2001; 2001US-269794P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Peder J, Lee L, Chen J, Jackson D, Ramenathan C, Siemens N;  
 PI Chang H, Carroll P;  
 NR WPI; 2002-691617/74.  
 New potassium channel beta-subunit, K+betaM2, proteins and nucleic  
 acids, useful for diagnosing, treating and/or preventing e.g.  
 reproductive, neural, metabolic, endocrine, memory, neurodegenerative  
 disorders or diseases -  
 XX  
 XX Example 1; Page 349-350; 366pp; English.  
 The present invention relates to human potassium channel beta-subunit  
 (K+betaM2) proteins and polynucleotides encoding such proteins. The  
 K+betaM2 sequences are useful for diagnosing, treating and/or preventing  
 reproductive disorders, neural disorders, disorders related to aberrant  
 potassium regulation or hyper potassium channel activity, metabolic  
 disorders (e.g. premature puberty), endocrine disorders (e.g. aberrant  
 growth hormone synthesis and/or secretion), memory disorders (e.g. aberrant  
 of the testis (e.g. spermatogenesis), neuroendocrine condition related  
 to aberrant thyroid hormone release, renal disease or disorders (e.g.  
 nephritis) disorders related to aberrant higher brain function (e.g.  
 learning deficiencies), neurodegenerative diseases (e.g. Alzheimer's  
 disease), proliferative disorders (e.g. carcinoid tumour) and disorders  
 involving excessive smooth muscle tone or excitability (e.g. asthma).  
 CC  
 CC The invention may be used to modulate haemostatic or thrombolytic activity, to  
 CC treat or prevent blood coagulation diseases or disorders, blood platelet

CC diseases, wounds, autoimmune diseases, disorders or conditions (e.g.  
 rheumatoid arthritis), allergic reactions (e.g. asthma), organ rejection  
 or graft-versus-host disease, and hyperproliferative diseases. K+betaM2  
 CC sequences are also used in gene therapy. The present sequence is human  
 CC BAC AC008652 exon used to isolate K+betaM2 cDNA. This sequence is used  
 CC in the exemplification of the invention.  
 XX  
 SQ Sequence 769 BP; 209 A; 180 C; 184 G; 196 T; 0 other;  
 Query Match 22.4%; Score 769; DB 24; Length 769;  
 Best Local Similarity 100.0%; Pred. No. 6,5e-143;  
 Matches 769; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 393 AGCGATTTTAAATAGTATGATTCATCTTCCCTTCTTACAGTATGATTCAGATTA 452  
 Db 1 AGCTATTTTAAATAGTATGATTCATCTTCCCTTCTTACAGTATGATTCAGATTA 60  
 Qy 453 AGCGATTCATCTTGAATGATTCATCTTAAATGATTCATCTTGAATGATTCATCTT 512  
 Db 61 AGCTATTCATCTTGAATGATTCATCTTAAATGATTCATCTTGAATGATTCATCTT 120  
 Qy 513 CATTGCTCTGATGAAACTGATGATTCATCTTAAATGATTCATCTTGAATGATTCATCTT 572  
 Db 121 CATTGCTCTGATGAAACTGATGATTCATCTTAAATGATTCATCTTGAATGATTCATCTT 180  
 Qy 573 CCAACTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 632  
 Db 181 CCAACTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 Qy 633 ATTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692  
 Db 241 ATTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 Qy 693 ACAGCGCTAATGATCTAGCAAGGATCTCAAGGATGATGATGATGATGATGATGATGAT 752  
 Db 301 ACAGCGCTAATGATCTAGCAAGGATCTCAAGGATGATGATGATGATGATGATGATGAT 360  
 Qy 753 TCTTGTCCGTTAATCTAGCAAGGATCTCAAGGATGATGATGATGATGATGATGATGAT 812  
 Db 361 TCTTGTCCGTTAATCTAGCAAGGATCTCAAGGATGATGATGATGATGATGATGATGAT 420  
 Qy 813 TTCCAGAAAAAGGAAAGTGAATCTTCCAGCTCCCAAGCTTGTCA 872  
 Db 421 TTCCAGAAAAAGGAAAGTGAATCTTCCAGCTCCCAAGCTTGTCA 480  
 Qy 873 AACTCTGACCCCTGATGAAATCAAGCAAGGATGATGATGATGATGATGATGATGAT 932  
 Db 481 AACTCTGACCCCTGATGAAATCAAGCAAGGATGATGATGATGATGATGATGATGAT 540  
 Qy 933 AAGATGCTCCCAAGGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 992  
 Db 541 AAGATGCTCCCAAGGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 Qy 993 ACCGAGTGGGTTTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1052  
 Db 601 ACCGAGTGGGTTTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 Qy 1053 GACAGGAGATGCAAGTCTGAGAGATGATGATGATGATGATGATGATGATGATGAT 1112  
 Db 661 GACAGGAGATGCAAGTCTGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 720  
 Qy 1113 TGCGAAGAAAGTCTTGAAGAACTTGAATGAAAGGAAAGCTTGA 1161  
 Db 721 TGCGAAGAAAGTCTTGAAGAACTTGAATGAAAGGAAAGCTTGA 769  
 RESULT 4  
 ABA09216/c  
 ID ABA09216 standard; cDNA; 906 BP.  
 AC ABA09216;  
 XX  
 XX 11-JAN-2002 (first entry)

XX DE Human VM106R.1 homologue-encoding cDNA, SEQ ID NO:992.  
 XX XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 XX haematopoietic regulation; tissue growth; immunomodulator; actin;  
 XX inhibin; chemotaxis; chemokines; thrombolytic; oncogenesis;  
 XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 XX chronic inflammatory condition; proliferative retinopathy;  
 XX atherosclerosis; coronary heart disease; arterial ischaemia;  
 XX bone disorder; osteoporosis; vascular growth disorder;  
 XX tissue regeneration; wound healing; infection; immune disorder;  
 XX cell culture; drug screening; gene therapy; antiinflammatory;  
 XX antineoplastic; antirheumatic; hemostatic; antiatherosclerotic;  
 XX cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 XX antitumoral; antiviral; antilucer; ss.  
 XX Homo sapiens.  
 PN MO200157188-A2.  
 XX 09-AUG-2001.  
 PD 05-FEB-2001; 2001MO-US03800.  
 PF 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT;  
 XX WPI: 2001-457740/49.  
 DR P-PSDB; ABB11972.  
 XX Human proteins and DNA encoding sequences useful for preventing  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX Claim 1; Page 844-845; 1963pp; English.  
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 XX invention also relates to vectors and recombinant host cells comprising a  
 XX nucleotide of the invention, methods of producing the novel polypeptides,  
 XX antibodies against the polypeptides, methods of detecting the nucleotides  
 XX or polypeptides in a sample, and methods of identifying compounds which  
 XX bind to polypeptides of the invention. Although novel, many of the  
 XX polypeptides of the invention have homology to known proteins, thereby  
 XX giving an insight into their probable biological activities, and hence  
 XX potential therapeutic applications. The polypeptides of the invention may  
 XX have various activities, including cytokine, cell proliferation or cell  
 XX differentiation activities; stem cell growth factor activity;  
 XX haematopoietic regulatory activity; tissue growth factor activity;  
 XX immunomodulatory activity; actin- or inhibin-related activities;  
 XX chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 XX thrombolytic activities; receptor or ligand activities; or may be  
 XX involved in oncogenesis, cancer cell proliferation or metastasis.  
 XX Depending on their biological activities, polypeptides and nucleotides of  
 XX the invention are useful for preventing, treating or ameliorating medical  
 XX conditions, e.g. by protein or gene therapy. Such conditions include  
 XX cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell  
 XX disorders), chronic inflammatory conditions (e.g. asthma or arthritis),  
 XX proliferative retinopathy, atherosclerosis, coronary heart disease,  
 XX arterial ischaemia, bone disorders (e.g. osteoporosis), and abnormal  
 XX vascular growth. Polypeptides involved with tissue regeneration and  
 XX repair (or nucleic acids encoding them) may be used to promote wound  
 XX healing (e.g. of burns, incisions and ulcers), while those with  
 XX immunomodulatory activities may be used in the treatment of viral,  
 XX bacterial and fungal infections in addition to immune disorders.  
 XX Polypeptides with growth factor activity may be used in cell cultures to  
 XX promote cell growth. For example, such polypeptides may be used to  
 XX manipulate stem cells in culture to give rise to neuroepithelial cells

CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a cDNA encoding a  
 CC novel human polypeptide of the invention.  
 XX SQ Sequence 906 BP; 220 A; 225 C; 216 G; 245 T; 0 other;  
 Query Match 20.2%; Score 699.2; DB 22; Length 906;  
 Best Local Similarity 98.9%; Pred. No. 4.6e-129;  
 Matches 704; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 515 ATGGCTGAGTGAAGAACTGTAGTCTTATATCTCGAAGAACAGAGGTCGCCAGTTCCC 574  
 DB 906 ATGGCTGAGTGAAGAACTGTAGTCTTATATCTCGAAGAACAGAGGTCGCCAGTTCCC 847  
 QY 575 AACTTCCTCCCTAGAGGTGTAGAGTGAATGTCGGGGGTCAAGTTATTTTACTGCGCAT 634  
 DB 846 AACTTCCTCCCTAGAGGTGTAGAGTGAATGTCGGGGGTCAAGTTATTTTACTGCGCAT 787  
 QY 635 TCACATTGATTAAGCATCCCTTCATTCCTCTCTGTGAAATGTTTCCCAAAGAGAC 694  
 DB 786 TCACATTGATTAAGCATCCCTTCATTCCTCTCTGTGAAATGTTTCCCAAAGAGAC 727  
 QY 695 ACGGCTAATGATCTACCCAAAGGACTCCAAAGGAAGTTTTCATTGACAGATGATTC 754  
 DB 726 ACGGCTAATGATCTACCCAAAGGACTCCAAAGGAAGTTTTCATTGACAGATGATTC 667  
 QY 755 TTGTTCCGTTATATTTGAGACTATCTCAGGAGACAGAGAGTGTCTCTGATCACTTT 814  
 DB 666 TTGTTCCGTTATATTTGAGACTATCTCAGGAGACAGAGAGTGTCTCTGATCACTTT 607  
 QY 815 CCAGAAAAAAGAAAGCTGAAAGAGGAAAGCTGAATACCTGCAGCTCCAGACTTGTGAA 874  
 DB 606 CCAGAAAAAAGAAAGCTGAAAGAGGAAAGCTGAATACCTGCAGCTCCAGACTTGTGAA 547  
 QY 875 CTCCTGACCCCGGATGAATCAAGCAAGCCGAGATGAATTCGACAGAGACTTGA 934  
 DB 546 CTCCTGACCCCGGATGAATCAAGCAAGCCGAGATGAATTCGACAGAGACTTGA 487  
 QY 935 GATGCTTCCCAAGAAAGCAACAAAGATTCGCCCCCTTCTCTCTGCTCCGCGGAC 994  
 DB 486 GATGCTTCCCAAGAAAGCAACAAAGATTCGCCCCCTTCTCTCTGCTCCGCGGAC 427  
 QY 995 GCGAAGTGGGTTTCATTAATCTGTGGTTACAGAGATCTGCACTTGGGCAAGAGGGA 1054  
 DB 426 GCGAAGTGGGTTTCATTAATCTGTGGTTACAGAGATCTGCACTTGGGCAAGAGGGA 367  
 QY 1055 CAGGAGATGCAAGATTTGCGAGAGTTCCCGGATTTTGGTTTGGAAAGATTTCCCTTG 1114  
 DB 366 CAGGAGATGCAAGATTTGCGAGAGTTCCCGGATTTTGGTTTGGAAAGATTTCCCTTG 307  
 QY 1115 GCAAAAGAGCTTTGAGAACTTTGAATGAAGACAGACCTTGATCGAGCCCAAGAA 1174  
 DB 306 GCAAAAGAGCTTTGAGAACTTTGAATGAAGACAGACCTTGATCGAGCCCAAGAA 247  
 QY 1175 AGATACCTCCGAGATTTTATCTCAATTCAGACAGCTGGAAGAGGCTTTTG 1226  
 DB 246 AGATACCTCCGAGATTTTATCTCAATTCAGACAGCTAATGGGGGCACTTG 195  
 RESULT 5  
 AAS34230  
 ID AAS34230 standard; cDNA; 440 BP.  
 XX AAS34230;  
 AC AAS34230;  
 XX 17-DEC-2001 (first entry)  
 DT Human cDNA encoding a novel foetal antigen, SEQ ID No 754.  
 XX Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;  
 XX immunomodulator; cardiovascular; cytoskeletal; nephrothropic;  
 KW



PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251889.  
 PR 08-DEC-2000; 2000US-0251890.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Rosen CA, Barash SC, Ruben SM,  
 XX MPI; 2001-488782/53.  
 DR P-PSDB; AAU21410.  
 XX  
 PT New polynucleotides and polypeptides for diagnosing, treating,  
 PT preventing or prognosing e.g. diseases or disorders of the nervous,  
 PT musculoskeletal, excretory, gastrointestinal, reproductive, and  
 PT respiratory systems -  
 Claim 1; SEQ ID No 754; 642bp; English.

XX The invention relates to novel nucleic acids encoding novel human foetal  
 CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.  
 CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. The antibodies to the antigens can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis, the  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular diseases e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiodysplasia, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular diseases e.g. corneal infection. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. Numerous  
 CC examples of diseases and disorders treated by the nucleic acids and  
 CC proteins are given in the specification. The present sequence

Query Match 12.2%; Score 423.4; DB 22; Length 440;  
 Best Local Similarity 98.9%; Pred No. 1.7e-74;  
 Matches 435; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

OY 1842 TTGGAATTAATCTCAATTAAGGATTTTAAAGGAAAAATCACTAATAT 1901  
 DB 1 TTGGAATTAATCTCTTAAGGATTTTAAAGGAAAAATCACTAATAT 60  
 OY 1902 GCACTTTCTTGAACCAATAGTCATTAATAGTACCTGCTTACTTAC 1961  
 DB 61 GCACTTTCTTGAACCAATAGTCATTAATAGTACCTGCTTACTTAC 120  
 OY 1962 CTAAACATGTAATCAACAGGATGATTTCTTCAATGTGAAGTACAGAAATCTT 2021  
 DB 121 CTAAACATGTAATCAACAGGATGATTTCTTCAATGTGAAGTACAGAAATCTT 180  
 OY 2022 TTTAGTATTTGTTTCTTACTTCTCCCATGTCATTAATCTTAATATATAGAG 2081  
 DB 181 TTTAGTATTTGTTTCTTACTTCTCCCATGTCATTAATCTTAATATATAGAG 240  
 OY 2082 CCAGTACGTAATAGTGTAGAGGCTTGGAGATCTTATATCCAACTGGG--TTTT 2140  
 DB 241 CCAGTACGTAATAGTGTAGAGGCTTGGAGATCTTATATCCAACTGGGTTTTT 300  
 OY 2141 TCTCATCTCTTACCTCCCTCTTGGATAGAGGTATGTTAATAAATATATGGCCCA 2200  
 DB 301 TCTCATCTCTTACCTCCCTCTTGGATAGAGGTATGTTAATAAATATATGGCCCA 360

OY 2201 ATGCAATTAATTTGAAATTTTAAATTTGCTTTCTTTGTTATAGGGGTGGGGCA 2260  
 DB 361 ATGCAATTAATTTGAAATTTTAAATTTGCTTTCTTTGTTATAGGGGTGGGGCA 420  
 OY 2261 ATGCAATTAATTAATTAATTTGCTTTCTTTGTTATAGGGGTGGGGCA 2320  
 DB 421 ATGCAATTAATTAATTAATTTGCTTTCTTTGTTATAGGGGTGGGGCA 480

RESULT 6  
 AAD49513  
 ID AAD49513 standard; cDNA; 2398 BP.  
 XX  
 AC AAD49513;  
 XX  
 DT 24-MAR-2003 (first entry)  
 XX  
 DE Human TRICH-15 cDNA.  
 XX  
 KM Human; transporter and ion channel; TRICH; atherosclerosis; cancer;  
 XX gene therapy; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH KEY Location/Qualifiers  
 FT CDS 114..1535  
 FT /tag= a  
 FT /product= "Human TRICH protein"  
 FT sig\_peptide 114..230  
 FT /tag= b  
 FT mac\_peptide 231..1532  
 FT /tag= c  
 FT /product= "Mature human TRICH protein"  
 WO200283712-A2.  
 XX  
 PN 24-OCT-2002.  
 XX  
 PF 12-APR-2002; 2002WO-US11760.  
 XX  
 PR 12-APR-2001; 2001US-283440P.  
 XX  
 PR 20-APR-2001; 2001US-285592P.  
 XX  
 PR 27-APR-2001; 2001US-287263P.  
 XX  
 PR 04-MAY-2001; 2001US-288666P.  
 XX  
 PR 18-MAY-2001; 2001US-292042P.  
 XX  
 PR 25-JAN-2001; 2001US-293724P.  
 XX  
 PR 22-JAN-2002; 2002US-351107P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Baughn MR, Elliott VS, Hafalia AJA, Yang J, Walla NK, Ramkumar J;  
 PI Forsythe TJ, Lu Y, Tang YT, Yue H, Raunam BE, Lai PG, Azimzai Y;  
 PI Lu DM, Gandhi AR, Thornton M, Nguyen DB, Arizazu CS, Emerling BM;  
 PI Svarnakar A, Yao MG, Ding L, He A, Griffin JA, Sanjanwalla MM;  
 PI Gietzen KD, Lee EA, Xu Y, Au-Young JK, Das D, Lee SY, Chang H;  
 XX WPI; 2003-092996/08.  
 DR P-PSDB; AAE32081.  
 XX  
 PT New human functional transporters and ion channels (TRICH)  
 PT polypeptides, useful for preparing a composition for diagnosing or  
 PT treating a disease associated with decreased expression or  
 PT overexpression of TRICH e.g. cancer -  
 Claim 5; Page 200-201; 204pp; English.

XX The invention relates to human transporters and ion channels (TRICH)  
 CC polypeptides and nucleic acid molecules encoding such polypeptides.  
 CC TRICH proteins are useful for preparing compositions for diagnosing or  
 CC treating diseases or conditions associated with decreased expression  
 CC or overexpression of functional TRICH e.g. atherosclerosis or cancer.  
 CC The invention is useful in gene therapy. The present sequence is  
 CC human TRICH cDNA.





XX 12-SEP-2002.  
 PD 21-FEB-2002; 2002WO-US05674.  
 XX 21-FEB-2001; 2001US-270132P.  
 XX 27-MAR-2001; 2001US-278953P.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Feder J, Lee L, Chen J, Jackson DG, Ramanathan C, Siemens N;  
 PI Chang H;  
 PI WPI: 2002-713455/77.  
 DR P-BSDb; ABU10886.  
 XX New polynucleotide encoding human potassium channel beta subunit  
 PT polypeptide, useful for diagnosing, preventing, treating or  
 PI ameliorating e.g. cancer.  
 XX  
 XX Claim 1; Fig 1; 332pp; English.  
 The invention relates to an isolated polynucleotide encoding a potassium  
 channel beta subunit (K-beta6) polypeptide or its variants. The human  
 potassium beta subunit polynucleotide or polypeptide is useful for  
 diagnosing, preventing, treating or ameliorating a pathological condition  
 such as gastrointestinal, reproductive, neural, sleep, cardiovascular or  
 CC pulmonary disorders, a disorder related to hyperpotassium channel  
 CC activity, an immune disorder related to aberrant NF-kB activity, pineal  
 CC gland associated disorders, migraine headaches, disorders associated with  
 CC aberrant melatonin synthesis and/or release or with low DNA repair  
 CC capacity or low free-radical buffering capacity, delayed sleep phase  
 CC syndrome, aberrations in circadian cycle, mammary cancer tumorigenesis,  
 CC age related disorders associated with decreased melatonin secretion, or  
 CC cancer. This polynucleotide sequence represents the cDNA encoding the  
 CC potassium channel beta subunit (K-beta6) protein of the invention.  
 XX  
 XX Sequence 2052 BP; 380 A; 640 C; 607 G; 425 T; 0 other:  
 SQ  
 Query Match 4.8%; Score 167; DB 24; Length 2052;  
 Best Local Similarity 64.6%; Pred. No. 1.5e-23;  
 Matches 267; Conservative 0; Mismatches 140; Indels 6; Gaps 1;  
 QY 967 CCCCCTTCCTCCTCCTCCTCCTGCGACCGCAAGTGGGTTTCATTACTGTGGTTACAG 1026  
 DB 705 CAGCGCGTCCAGTGCCTGAGACGCGACCGCGCTCGGCTACATCAATCGGCTACCG 764  
 QY 1027 AGATCTGCACCTTGGGAG 1086  
 DB 765 CGGCTCTACACCATCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 824  
 QY 1087 GATTGGTTTGGAGAGATTCTCTTGGCAAAAGAGCTTTGAGAACTTTGAAATGA 1146  
 DB 825 CATCACCGTTTGGAGAAAGAGTCTGCGCAAGAGAGTGTGGGACACCGTGAACGA 884  
 QY 1147 AAGCAGAGACCTGATGAGAGCGGCAAAAGATCACTTCAAGTTTATTCGAATTCGA 1206  
 DB 885 AAGCGGAGACCGGACCTCCCGAGCGCTCACTGCGCTTATACCTCAAGTTTCA 944  
 QY 1207 GCACCTGAAAAGGCTTTTGAATGTTTGAAGTGTGATTCACATGTGGCTGTGA 1266  
 DB 945 CTTCCTGAGACGAGGCTTTCGACCAAGCTGTCCGAGTGGGCTTCCACATGAGGCTGACG 1004  
 QY 1267 CTATCGGTGACAGATCTTT-----CATCAACCAATATACAGATGACAGATCTGTC 1320  
 DB 1005 CTCACCGGACCTCGGCTTTCGACGACGACGACGACGACGACGACGACGACGACGAC 1064  
 QY 1321 AAGCTACACTGAATATGTTCTTCTACCGTGAAGCTTCAAGATGTAACCTGAC 1373  
 DB 1065 CAGCTACACGAGTATGTTCTTCTGACAGGAGTGAAGCTCCCGACGACCCCTGCG 1117  
 RESULT 10

ABO88125/c  
 ID ABO88125 standard; cDNA; 109201 BP.  
 XX  
 AC ABO88125;  
 XX  
 DT 18-SEP-2002 (first entry)  
 XX  
 DE Human osteoblast differentiation related cDNA SEQ ID NO 32.  
 XX  
 KW Human; osteoblast; stem cell differentiation; bone tissue deposition;  
 XX osteoporosis; osteopathic; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200250301-A2.  
 XX  
 PD 27-JUN-2002.  
 XX  
 PF 18-DEC-2001; 2001WO-US48276.  
 XX  
 PR 18-DEC-2000; 2000US-255882P.  
 XX  
 PR 24-APR-2001; 2001US-285691P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX (PROC ) PROCTER & GAMBLE CO.  
 PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
 PI Mertz L;  
 PI WPI: 2002-557663/59.  
 XX  
 XX The invention relates to genes and their expression profiles are used  
 XX for:  
 CC (a) screening modulators of precursor stem cell differentiation into  
 CC osteoblasts, or bone tissue deposition;  
 CC (b) diagnosing abnormal deposition of bone tissue, abnormal rate of  
 CC osteoblast formation or osteoporosis; or  
 CC (c) treating or monitoring treatment of the conditions cited in (b), or  
 CC monitoring the progression of bone tissue deposition.  
 CC Specific conditions include postmenopausal osteoporosis, glucocorticoid  
 CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy,  
 CC drug-induced abnormalities in bone formation or bone loss, conditions  
 CC that involve altered bone metabolism (e.g. idiopathic juvenile  
 CC osteoporosis), skeletal disease linked to breast cancer, mastocytosis,  
 CC Fanconi syndrome or fibrous dysplasia. The present sequence is that of an  
 CC osteoblast differentiation associated cDNA marker of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 XX Sequence 109201 BP; 32871 A; 23488 C; 22108 G; 30734 T; 0 other:  
 SQ  
 Query Match 4.8%; Score 167; DB 24; Length 109201;  
 Best Local Similarity 64.6%; Pred. No. 3.5e-23;  
 Matches 267; Conservative 0; Mismatches 140; Indels 6; Gaps 1;  
 QY 967 CCCCCTTCCTCCTCCTCCTCCTGCGACCGCAAGTGGGTTTCATTACTGTGGTTACAG 1026  
 DB 9291 CAGCGCGTCCAGTGCCTGAGACGCGACCGCGCTCGGCTACATCAATCGGCTACCG 9232  
 QY 1027 AGATCTGCACCTTGGGAG 1086  
 DB 9291 CGGCTCTACACCATCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 9172  
 QY 1087 GATTGGTTTGGAGAGATTCTCTTGGCAAAAGAGCTTTGAGAACTTTGAAATGA 1146

Db 9171 CATCACGTTTGGCGAAGACGTGCGCCAGAGGTTGTTGGGAGAACCTTGAACGA 9112  
 Qy 1147 AAGCAGAGACCTGATGAGCCCGCAGAAAGATACCTCCGAGATTTATCTCAATTTCAA 1206  
 Db 9111 AGCCGGGAGCCCGACCGTCCCGAGGCGTACACCTCGCGCTATTTACTCAAGTTTCAA 9052  
 Qy 1207 GCACCTGGAAGGCTTTGATGTTGTCAGAGTGGAGATTCACATGATGATGCTGTA 1266  
 Db 9051 CTTCTGAGACAGGCTTTCAGACGCTGTCGAGTCGGGCTTCCATGCTGCTGAG 8992  
 Qy 1267 CTCATCGGTGACAGATCTTT-----CATCAACATATACAGATGACAAATCTGCTC 1320  
 Db 8991 CTCACGGGACCTGCGCTTTCGACAGCACGACGACGAGGAGAGACAAAGATCTGAGAC 8932  
 Qy 1321 AAGCTACACTGAATATATGTCTTCTACCGTGAAGCTTTCAGATGTCACCTCAAC 1373  
 Db 8931 CAGCTACACCGAGTACGTCCTTCTGACAGGAGTGAAGCTCCCGACACCTCCCTCCG 8879  
 RESULT 11  
 ID ABO40654 standard; DNA; 854 BP.  
 AC ABO40654;  
 DT 12-JUL-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27245.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 OS Homo sapiens.  
 PN WO200218632-A2.  
 PD 07-MAR-2002.  
 PF 01-SEP-2001; 2001WO-EP10074.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 Y (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX WPI; 2002-371829/40.  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the

CC methylation status of many C residues to be determined simultaneously.  
 CC ABO41410-ABO41412 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SO Sequence 854 BP; 131 A; 98 C; 289 G; 336 T; 0 other;  
 Query Match 3.3%; Score 114.2; DB 24; Length 854;  
 Best Local Similarity 59.7%; Pred. No. 3.6e-13;  
 Matches 213; Conservative 0; Mismatches 138; Indels 6; Gaps 1;  
 Qy 1001 TCGGGTTTCATTACTGCTGCTTACAGAGATCTCGACCTTGGCAGAGAGGACAGCA 1060  
 Db 491 TCGGGTAT 550  
 Qy 1061 GATGCCAAGTTTCGAGAGATTCGCCGAGTTTGGTTGTGGAAGATTTCTTGGCAAA 1120  
 Db 551 GACGTTAAGTTTCGCGAGAGTGGCGGCTATTTGTTGCGAAGAGCTGTTGGTTAAG 610  
 Qy 1121 GAAGTCTTTGGAGAACTTTGAAATGAAGACAGACACCTCATGACGCCCGAAGAAATAC 1180  
 Db 611 GAGGTGTTGGGATATTTTGAACGAAAGTCCGAGTTTCATGCTTTTCGAGAGCTTAT 670  
 Qy 1181 ACCTCAGATTTTATCTCAATTCAGACACCTGGAAGGCTTTGTATATGTGTCAAG 1240  
 Db 671 ATTCGCGTAT 730  
 Qy 1241 TGTGATTCACATGTTGGCTGTATCATGCTCATGCTGACAGATCTTT-----CATCAAC 1294  
 Db 731 TCGGGTTTATATATGTTGGCGGTATTTTACGGGTATTTGCGTTTGTATGATATC 790  
 Qy 1295 CAAATACAGATGACAGATCTGCTCAAGCTCACTGAAATATGCTTCAACGTGAG 1351  
 Db 791 GATTAGACGAGATTAAGATTTGATTAATATATGATTAATATATGATGATGTTTGTAGGAG 847  
 RESULT 12  
 ID ABO40655/c  
 AC ABO40655 standard; DNA; 854 BP.  
 DT 12-JUL-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27246.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 OS Homo sapiens.  
 PN WO200218632-A2.  
 PD 07-MAR-2002.  
 PF 01-SEP-2001; 2001WO-EP10074.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 Y (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX WPI; 2002-371829/40.  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.





XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 260.  
 DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX Homo sapiens.  
 OS  
 XX MO200218632-A2.  
 PN  
 XX 07-MAR-2002.  
 PD  
 XX 01-SEP-2001; 2001WO-EP10074.  
 PF  
 XX 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 (EPIG-) EPIGENOMICS AG.  
 PA Olek A, Piepenbrock C, Berlin K, Guefig D;  
 PI WPI; 2002-371829/40.  
 DR  
 XX Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 XX amplicons from chemically treated DNA  
 XX  
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
 PS  
 XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridized to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridization to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridized to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations of single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB01310-AB054121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 CC  
 XX  
 SQ Sequence 1757 BP; 662 A; 640 C; 209 G; 246 T; 0 other;  
 Query Match 3.2%; Score 109.8; DB 24; Length 1757;  
 Best Local Similarity 62.6%; Pred. No. 3,1e-12;  
 Matches 171; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
 Oy 1001 TGGGGTTTCATTACGTGGTTCAGAGATCCGACCTTGGCCAGAGAGGACAGCA 1060  
 Db |||||  
 Oy 273 TCGGTTATATTATATTCGTTATCGCGGTTTATATTCGCGGAGACGGTGGCG 214  
 Db |||||  
 Oy 1061 GATGCGAATTCGAGAGTCCCGGATTTGGTTGGGAAGATTCTCTTGGGAAA 1120  
 Db |||||  
 Oy 213 GACGTTAAGTTTCGGCGAGTGGCCGATTAATTCGTTGGGAAAGAGCGTGGTTAG 154  
 Db |||||  
 Oy 1121 GAAGCTTTGGAAGAACTTGAATGAAGACAGACCCGATCCAGCCCAAGAAATAC 1180  
 Db |||||  
 Oy 153 GAGGTGTTGGGAAATTTTGAACGAAGTCGAGATTTCGATCTTTTGGAGCGTTAT 94  
 Db |||||  
 Oy 1181 ACCTCGAATTTATCTCAATTAAGACACCTGGAAGGGCTTTGATATGTTGTCAAG 1240  
 Db |||||  
 Oy 93 ATTTCGGTTATTTATTTAAGTTTATTTTGGAGAGGTTTTCGATGAGTTGCGAG 34  
 Db |||||

Oy 1241 TGTGATTCCACATGCTGGCTTAACTCATCG 1273  
 Db |||||  
 Oy 33 TCGGGTTTATATAGTGGCGGTATGTTTACG 1  
 Db |||||  
 RESULT 15  
 ABQ40656/c  
 ID ABQ40656 standard; DNA; 854 BP.  
 XX  
 AC ABQ40656;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27247.  
 XX  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX MO200218632-A2.  
 PN  
 XX 07-MAR-2002.  
 PD  
 XX 01-SEP-2001; 2001WO-EP10074.  
 PF  
 XX 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 (EPIG-) EPIGENOMICS AG.  
 PA Olek A, Piepenbrock C, Berlin K, Guefig D;  
 PI WPI; 2002-371829/40.  
 DR  
 XX Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 XX amplicons from chemically treated DNA  
 XX  
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
 PS  
 XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridized to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridization to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridized to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations of single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB01310-AB054121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 CC  
 XX  
 SQ Sequence 854 BP; 132 A; 98 C; 302 G; 322 T; 0 other;  
 Query Match 3.1%; Score 108.6; DB 24; Length 854;  
 Best Local Similarity 57.1%; Pred. No. 4,6e-12;  
 Matches 220; Conservative 0; Mismatches 159; Indels 6; Gaps 1;  
 Oy 967 CCCCCCTTCTCCCTGCTCCCTGCGCAGCGGAGTTCATTACTGTGGTTACG 1026  
 Db |||||

```
Db 338 CACGCCCTCCCAATCGCTAAACGACACCGAGCTCGAATCATCACCATCGACTACCG 339
Qy 1027 AGATTCCTGACCTTGGGAGAGGAGACAGCAGATGCCAAGTTTGGAGAGTTCCCG 1086
Db 338 CGACTCCTACCATCGAACGAAACGGCGAAACGAAACGCCAAATTCGACGATTAAGCG 279
Qy 1087 GATTTGGTTTGGAGAGATTCTCTGGCAAAAGAGCTTTGGAGAACTTTGAATGA 1146
Db 278 CATCACCGTTTAGAABAAAACGTCCCTAACCAAAAATTTAAAAACACCCTAAACGA 219
Qy 1147 AAGCAGAGACCTGATCGAGCGCCGAAAGATACCTGCAATTTATCTCAATTCAA 1206
Db 218 AAACCGAAACCCGACCGTCCCGAAGCTACACCTCGCGCTATTACCTGAAATTGAA 159
Qy 1207 GCACCTGAAAGGCGTTTGTATATGTGTCAAGTGTGATTCACATGTGGCGCTGAA 1266
Db 158 CTTCCTTAAACAAACCTTGCACAACTATCCGAATCGAACTTCACATATTAACGTACAA 99
Qy 1267 CTGATGGTGACAGCATCTTT-----CATCAACCAATATACAGATGACAAGATCTGTC 1320
Db 98 CTCACGAAACCTTACGCTTACCAACACACGACCAAAACGAAACAAATCTAAC 39
Qy 1321 AAGCTACACTGAATATGTCTTCTAC 1345
Db 38 CAACTACACCGAATACGTCTCTAC 14
```

Search completed: September 5, 2003, 15:10:11  
Job time : 850 secs